Establishing a Survey of Refugees in Germany: Challenges in Sampling, Field Work and Measurement

Dissertation

zur Erlangung des akademischen Grades

doctor philosophiae
(Dr. phil.)

im Fach Soziologie

Eingereicht am 24.04.2020

an der Kultur-, Sozial- und Bildungswissenschaftlichen Fakultät der Humboldt-Universität zu Berlin

von Herrn Jannes Jacobsen

Verteidigt am: 31.08.2020

Prof. Dr.-Ing. Dr. Sabine Kunst	Prof. Dr. Christian Kassung
Präsidentin der Humboldt-Universität zu Berlin	Dekan der Kultur-, Sozial- und Bildungswissenschaftlichen Fakultät

Gutachter:

1. Prof. Dr. Jürgen Schupp, Freie Universität Berlin

2. Prof. Dr. Johannes Giesecke, Humboldt-Universität zu Berlin

Acknowledgments

I want to begin my acknowledgements by pointing to the one person who has guided me since I started my Master's at Freie Universität Berlin. Jürgen Schupp not only introduced me into the benefits of working with panel data but also showed me the way into academia. Were it not for Jürgen and his dedicated work, neither would I have become a PhD candidate at Humboldt University and the Socio-Economic Panel, nor would the IAB-BAMF-SOEP Survey of Refugees exist; and without this data, I could not have written this thesis. Second, I would like to thank Johannes Giesecke, my second supervisor, who has provided me with helpful guidance during the work on my manuscript and who always made me feel comfortable asking for help. Important to mention are my (former) colleagues at the Socio-economic Panel: Martin Kroh, Simon Kühne and Rainer Siegers introduced me into the bittersweet world of survey methodology, especially into sampling and weighting. Additionally, Martin and Simon have been wonderful co-authors. Magdalena Krieger, Lea-Maria Löbel, Jana Nebelin, and Andreas Franken have been there, every day, for almost four years, reading my manuscripts, reviewing my Stata code and providing terrific feedback on work in progress. Additionally, I would like to thank Lisa Pagel, Katja Schmidt, Luise Burkhardt, Philipp Eisnecker, Knut Wenzig, Valeriia Heidemann, and David Richter for creating a working environment in which coming to work every day is not a burden but something I am looking forward to. Moreover, I would like to thank Lukas Marian Fuchs who has been a terrific co-author for one of the articles. Additionally, I would like to thank Zerrin Salikutluk and Ferdinand Geißler for agreeing to being part of my PhD commission and thus sharing their expertise. I also would like to thank Deborah Bowen who has done a wonderful job in proofreading my manuscripts.

I would like to express my gratitude to Theresa Huber the one person who has accompanied me through my Bachelor's, my Master's and now my doctorate. Part of my achievements are due to her dedication in sharing her expertise in qualitative research and reminding me that there is a world outside quantitative research. Moreover, I thank Matthes Hackbusch for patiently explaining statistics to me over and over again.

As a sociologist, I am aware of path dependencies in educational trajectories. Therefore, I would like to mention my parents Sabine and Matthias who have done a wonderful job in providing me with skills and traits necessary in the academic world.

Contents

Summary

This dissertation looks at the different steps in the process of conducting a survey on refugees living in Germany and discusses key focal points of integration research. In four different articles, I discuss the novel sampling strategy used in a survey of refugees, analyze the effects of missing questionnaire languages, test latent constructs for measurement invariance, and discuss the analysis of economic integration in a changing migration regime.

In the first article, I propose a sequential sampling strategy to sample refugees in times of high immigration. This is necessary because sampling frames (e.g., official records) of immigrants usually have a time lag in covering the population of interest.

In the second article, I show that the lack of questionnaires in a respondent's mother tongue increases item nonresponse. Providing additional audio recordings of the questions does not diminish this effect. I find that due to this issue, item nonresponse is in many cases not missing completely at random and thus calls for further solutions in data analysis. In light of the vast amount of information covered in social surveys, I propose to correct for such item nonresponse by means of weighting adjustments.

In the third article, I use conceptions of democracy as a case study to show that latent constructs in multi-cultural and multi-linguistic surveys face specific challenges and limitations in their comparability. By employing tests for measurement invariance, my results show that conceptions of democracy are likely not comparable across countries of origin or across languages.

The fourth and last article looks at the economic integration of refugees. In it, I propose that integration trajectories have to be observed within the specific institutional settings in which they take place. Fixed-effects regression analyses combined with a coarsened exact matching lead to the conclusion that a secure residence permit and participation in integration classes lead to increasing investments in future labor market access of refugees in Germany.

In these four articles, I discuss some of the obstacles that emerge when implementing a survey of refugees for the purpose of research on refugee integration and provide solutions for handling them.

Zusammenfassung

Diese Dissertation betrachtet verschiedene Schritte einer sozialwissenschaftlichen Erhebung zur Integration Geflüchteter in Deutschland. Anhand von vier Zeitschriftenartikeln wird eine neuartige Strategie, um eine Zufallsstichprobe von Geflüchteten in Deutschland zu ziehen, besprochen, die Folgen fehlender muttersprachlicher Übersetzungen von Fragebögen analysiert, latente Konstrukte auf Vergleichbarkeit getestet und Fragen ökonomischer Integration in sich verändernden Migrationsregimen diskutiert.

Der erste Artikel befasst sich mit einer sequentiellen Ziehungsstrategie für Zufallsstichproben. Diese ermöglicht eine zeitnahe Erhebung von Zuwanderern in Zeiten hoher Immigration, da Registerdaten Migranten nur mit zeitlicher Verzögerung umfassend abdecken.

Im zweiten Artikel wird gezeigt, dass fehlende muttersprachliche Übersetzungen von Umfragen die Item-Nonresponse erhöhen. Auch die Bereitstellung von Audio-Aufnahmen kann diesem Effekt nicht entgegenwirken. Darüber hinaus wird deutlich, dass es sich hierbei in vielen Fällen um einen systematischen Effekt auf die Item-Nonresponse handelt, weshalb weiterführende Analysen einer statistischen Korrektur bedürfen. Da Befragungen in der Regel umfassende Informationen erheben, schlage ich vor, Item-Nonresponse-Gewichte zu schätzen.

Im dritten Artikel wird die Vergleichbarkeit latenter Konstrukte in multikulturellen und multisprachlichen Erhebungen am Beispiel von Vorstellungen zu demokratischen Systemen untersucht. Messinvarianztests deuten darauf hin, dass Vorstellungen von Demokratie über verschiedene Herkunftsländer und Sprachen nicht vergleichbar sind.

Der letzte Artikel beschäftigt sich mit der ökonomischen Integration Geflüchteter und argumentiert, dass diese auf institutioneller Ebene betrachtet werden muss. Fixed-Effects-Regressionsanalysen kombiniert mit einem exakten Matching führen zu der Schlussfolgerung, dass sichere Aufenthaltstitel und die Teilnahme an Integrationskursen bei Geflüchteten in Deutschland zu einer erhöhten Anstrengung führt Zugang zum Arbeitsmarkt zu bekommen.

Anhand dieser vier Artikel werden unterschiedliche Aspekte einer sozialwissenschaftlichen Erhebung hinsichtlich der Messung von Integration bei Geflüchteten in Deutschland diskutiert.

1 Introduction

During the years 2013-2016, over one million refugees migrated to Germany (BMI, 2016). Especially the summer and fall of 2015 proved to be historic. The expectation was that such a large number of refugees entering the country over such a short period of time would have major impacts on the social welfare system, the housing and labor markets, and on social cohesion (for an overview regarding ealier cohorts see Bloch & Levy, 1999; for a general overview see Münkler & Münkler, 2016; for a current debate see Pries, 2016). However, there was little knowledge about this cohort of refugees to back up these assumptions. As Germany has no publicly available register data on refugees containing more than basic administrative information, there was an urgent need for a survey on refugees that could allow for robust analysis and timely policy advice and social reporting. The IAB-BAMF-SOEP Survey of Refugees in Germany was launched in late 2015 to fill this gap (Brücker, Rother, & Schupp, 2017; Kühne, Jacobsen, & Kroh, 2019). This thesis follows the development of this data set by exploring the challenges faced during sampling, fieldwork, measurement, and analysis of integration trajectories of refugees.

This thesis is structured as follows. In the introduction, Chapter 1, I discuss the differing usages of the term "refugee" in the literature and the definition used here, I give insights into the data frame on which this thesis is based and I provide a general overview on the different chapters in this thesis. In Chapter 2, I discuss the research question of this thesis in detail. In Chapter 3, I review the existing research on migrant integration and cross-cultural survey methods and highlight the research gaps this dissertation aims to fill. In the articles that comprise Chapters 4-7, I deal with key aspects of the research question. Chapter 8 provides a conclusion and an outlook to future research that could extend the scope of this thesis.

1.1 The Subject: Refugees

This thesis is built around research on refugees. As the term refugee is used in different ways in the academic literature and depends heavily on the legal context, I will provide a brief definition in the following.

When talking about refugees in Germany, I refer to four different categories:

1) *Asylum seeker*. A person who migrated to Germany and applied for asylum, which is still under review.

2) *Refugees* according to the Geneva Convention on the Status of Refugees or According to the German *Grundgesetz* Art. 16a.
   According to the Geneva Convention, a refugee:

   > "is someone who is unable or unwilling to return to their country of origin owing to a well-founded fear of being persecuted for reasons of race, religion, nationality, membership of a particular social group, or political opinion." (UNHCR, 1951, p. 3)

   According to the *Grundgesetz* Art. 16a, people who are persecuted politically are entitled to refuge and thus fall under the safeguard of the Geneva Convention on the Status of Refugees.
3) People with *subsidiary protection.*
   People with subsidiary protection do not fulfill the criteria of refugees (according to the Convention or according to *Grundgesetz* Art. 16a), but the conditions in their home country pose a potential harm to their life. For example, people from countries experiencing (civil) war are entitled to this kind of protection.
4) People with a *suspension from deportation.*
   People with a suspension from deportation have applied for asylum, but their claim has been rejected. However, due to medical conditions or missing passports (or other administrative reasons) they must not be deported to their home country.

1.2 The Data Frame: The IAB-BAMF-SOEP Survey of Refugees

The idea for writing this thesis arose in the process of development and implementation of the IAB-BAMF-SOEP Survey of Refugees. I was lucky to be part of the research consortium of the Institute for Employment Research (IAB), the Federal Office for Migration and Refugees (BAMF), and the German Socio-Economic Panel (SOEP), which launched this study on refugees. My work in this project allowed me to follow the entire process of planning, sampling, and implementing the first waves of this survey of refugees in Germany. Therefore, this thesis takes a cutting-edge project as an example and demonstrates the importance of examining the different steps from data collection to data analysis when carrying out research on the integration of refugees.

The IAB-BAMF-SOEP Survey of Refugees is a random sample drawn from the central register of foreigners in Germany (AZR) and went into the field in summer of 2016. It is designed as a panel study, running for at least six years until 2022. The target population is made up of people that moved to Germany in the years 2013-2016 and who applied for asylum, irrespective

of the outcome. Up to 2020, the consortium successfully released three waves of data. These data serve as a database for academic research, social reports, and policy advice. The research group integrated the survey into the German Socio-economic Panel (SOEP), a German household panel study that has been running since 1984 (Göbel et al., 2019). As part of the SOEP, the IAB-BAMF-SOEP Survey of Refugees started in 2016 with 3,289 households and 4,465 individual interviews (SOEP sample M3 and M4)[1]. In 2017, a supplementary sample was added (SOEP sample M5, Jacobsen et al., 2019), bringing the size to 3,822 households and 5,595 individual interviews in 2017. In 2018, 3,061 households and 4,376 individual household members remained (unweighted). The response rates for the baseline waves were around 50%, whereas response rates increased to 66% in the second wave and 64% in the third wave. Design and nonresponse weighting as well as post-stratification are used to counteract a disproportionate sampling design, first-wave nonresponse, and panel attrition. The data are free of charge and distributed yearly as a scientific use file by the research data centers of the SOEP (SOEP, 2019) and the Institute for Employment Research (IAB, 2019). Table 1 displays weighted key characteristics of the survey in its latest, 2018, wave.

In order to fully understand the utility, importance, and indeed necessity of this refugee survey, one has to understand the scope of the SOEP in general. Its basic aim is to provide panel data on the living conditions of individuals and households in Germany (see Hanefeld, 1987 for a detailed overview). The target population comprises private households in Germany. The SOEP was launched specifically to learn about the development and distribution of income and net worth in Germany, labor market participation, changes in people's social environments and household composition, and subjective well-being. The SOEP serves as a cross-sectional and longitudinal database for social reporting and academic research and is used, for instance, by the German federal government in its report on poverty and wealth (*Armuts- und Reichtumsbericht*) and in analyses of the implementation of the 2014 minimum wage reform. Besides serving as a data source, the SOEP is also used to improve and enhance survey methodological techniques used in sampling, data collection, and measurement. These efforts help to ensure the consistently high quality of the SOEP data and to increase the reliability of social reporting and research based on the SOEP. Since the first wave of the SOEP in 1984 until its latest wave in 2018, the SOEP has interviewed 129,662 distinct households with 142,308 individuals. In its most recent, 2018, wave (v.35), the SOEP consists of 30,306 adult respondents in 18,754 households. These many panel waves resulting in multiple interviews per

[1] The name giving (M3, M4) results from the SOEP standard procedure to name sub-samples in alphabetical order. Prior to the IAB-BAMF-SOEP Survey of Refugees, samples M1 and M2 have been added, covering preceding migrant cohorts (see also figure 4 on page 13).

household and the high number of respondents in the cross-section are a unique data source for acquiring in-depth information on households and individuals in Germany.

Table 1: Key Characteristics of the IAB-BAMF-SOEP Survey of Refugees in 2018

	Variables	Share in % (weighted)
Individual Level	**Gender**	
	Male	69.6
	Female	30.4
	Age	
	18-29	51.3
	30-49	41.0
	50-65	6.7
	66+	1.0
	Country of Origin	
	Syria	44.4
	Afghanistan	13.7
	Iraq	10.5
	Eritrea	5.6
	Somalia	1.8
	Other	24.1
	Employment Status	
	Employed (including vocational training and internships)	35.5
	Not Employed	60.8
	Retired	0.2
	N.A.[1]	3.4
	Family Status	
	Married	36.3
	Single	47.5
	Divorced/Widowed	15.5
	N.A.[1]	0.7
	Asylum Status	
	Refugee (including Subsidiary Protection)	96.5
	Pending	16.0
	None (including Suspension of Deportation)	9.0
	Other (i.e. no refugee related status)	2.7
	N.A.	2.8
Household Level	**Number of Kids in Household**	
	0	65.7
	1	11.2
	2	10.3
	3+	12.8
	Inhabitants in Local Administrative Unit 2 (LAU2)[2]	
	Up to 5k	5.8
	Up to 50k	44.1
	Up to 500k	33.7
	More than 500k	16.4
	N(Individual)	4,351
	N(Household)	3,042

[1] N.A. refers to nonresponse due to refusal, not knowing the answer or insufficient provision of information for generating a variable (e.g. CASMIN).

[2] In Germany LAU2 refers to *Gemeinden* (see Eurostat, 2018, for further information).

In order to fulfill its function in providing a deeper understanding of developments in German society over time, the SOEP constantly needs to react to changes in the underlying target population and find survey methodological ways to do that. Changes in the underlying target population can either come from within (death, birth, emigration) or from outside (immigration). The former do not pose a substantial problem for SOEP data quality, as in theory, the panel study reflects all changes in the target population including fertility and mortality.

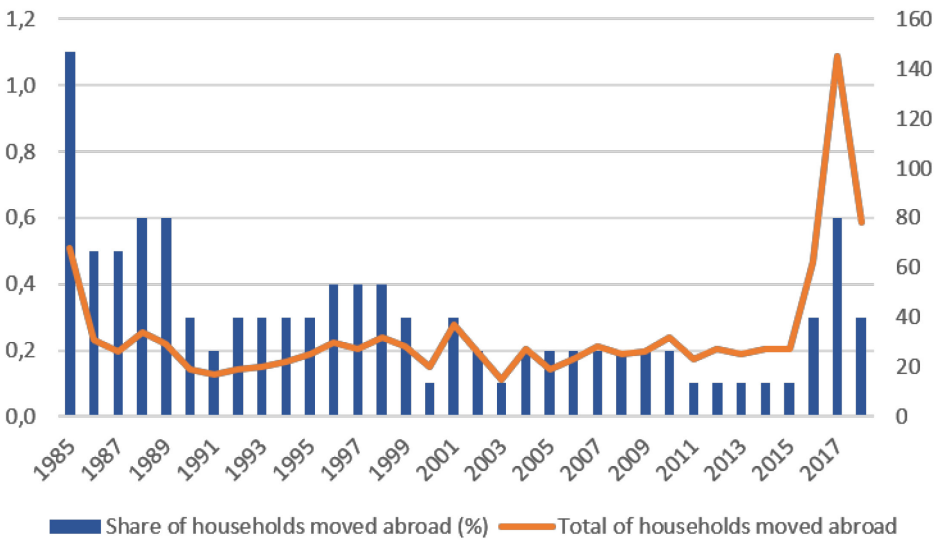

Figure 1: Emigrated SOEP households
Own Calculations based on N = 129,662 households, SOEP v.35 (SOEP, 2019)

As a consequence, the SOEP is a unique data source to study endogenous changes in German society. For instances, regarding integration and migration research, the SOEP can be used for emigration related analyses and research on "circular migration" (Dustmann & Kirchkamp, 2002) as panel attrition may occur due to moving abroad (e.g. Liebau & Schupp, 2011; Wagner, Schupp, & Rendtel, 1994, p. 96). Figure 1 indicates that panel attrition due to moving abroad has always existed in the SOEP, subsequently providing unique research potential. Pooling all SOEP waves until 2018 results in 1119 emigrated households (not displayed as a table) – a sufficient number for analyses. Immigration, however, is exogenous and thus calls for solutions, one of which is to add enlargement samples (Rendtel, 1995).

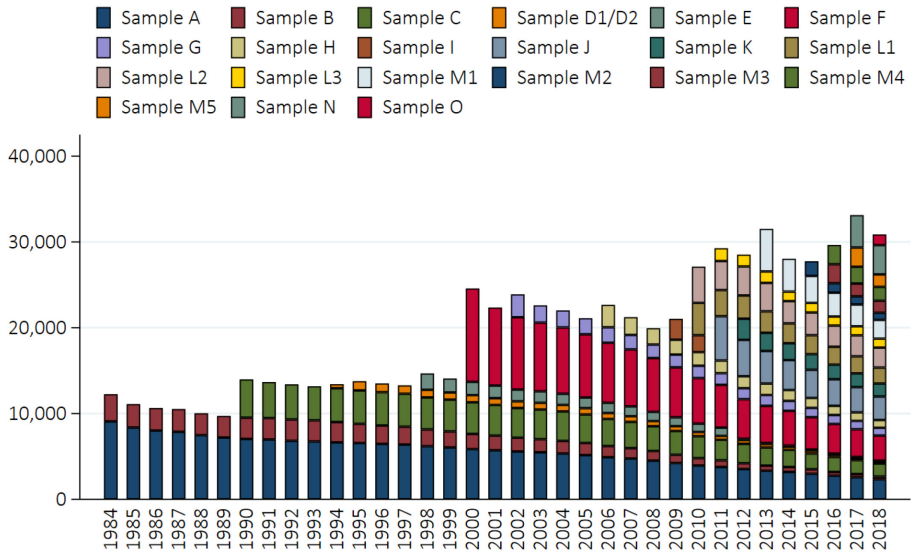

Figure 2: SOEP sub-samples over time and number of respondents
Own Calculations based on N = 142,308 individuals, SOEP v.35 (SOEP, 2019)

In order to react to changes in the target population and for allowing robust analyses (e.g., by providing sufficient sample size for sub-groups), the SOEP does not only comprise one general population survey but is the sum of a varied set of different sub-samples added over the years (see Figure 2). Per definition, all those samples comprise migrants, as migrants are part of the population in Germany (see Figure 3 for an aggregate share per sample over all survey years).

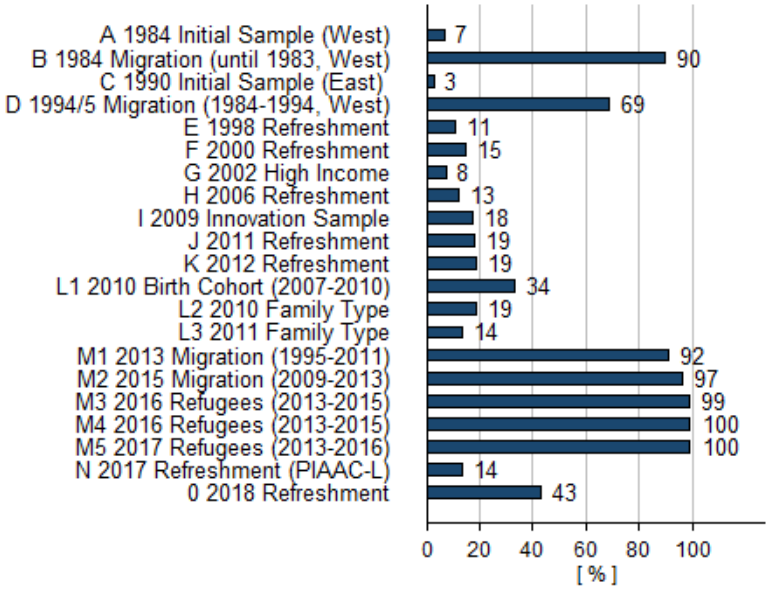

Figure 3: Aggregate share (over all survey years) of Respondents with Migration Background for each SOEP sample
Own Calculations based on N = 142,308 individuals, SOEP v.35 (SOEP, 2019)

However, in order to allow robust analyses with sufficient respondents, it is of interest to add extra enlargement samples for the purpose of immigration research. Therefore, as a reaction to

exogenous growth (i.e. immigration), the SOEP has added several immigration related enlargement samples so far (e.g. samples B, D, M1, M2). Other samples focus on specific sub-groups (e.g. high net worth individuals (sample G)) that are not sufficiently represented in general population surveys, and some samples are implemented to counteract panel attrition (such as samples J and K, (Kroh, Käppner, & Kühne, 2014).

In light of all these efforts to cover both the base population and important sub-groups in sufficient number, the SOEP provides a mirror of changes in the underlying target population of German households and thus tells the story of immigration to Germany over decades (see figure 4 for gross migration to Germany since the Second World War).

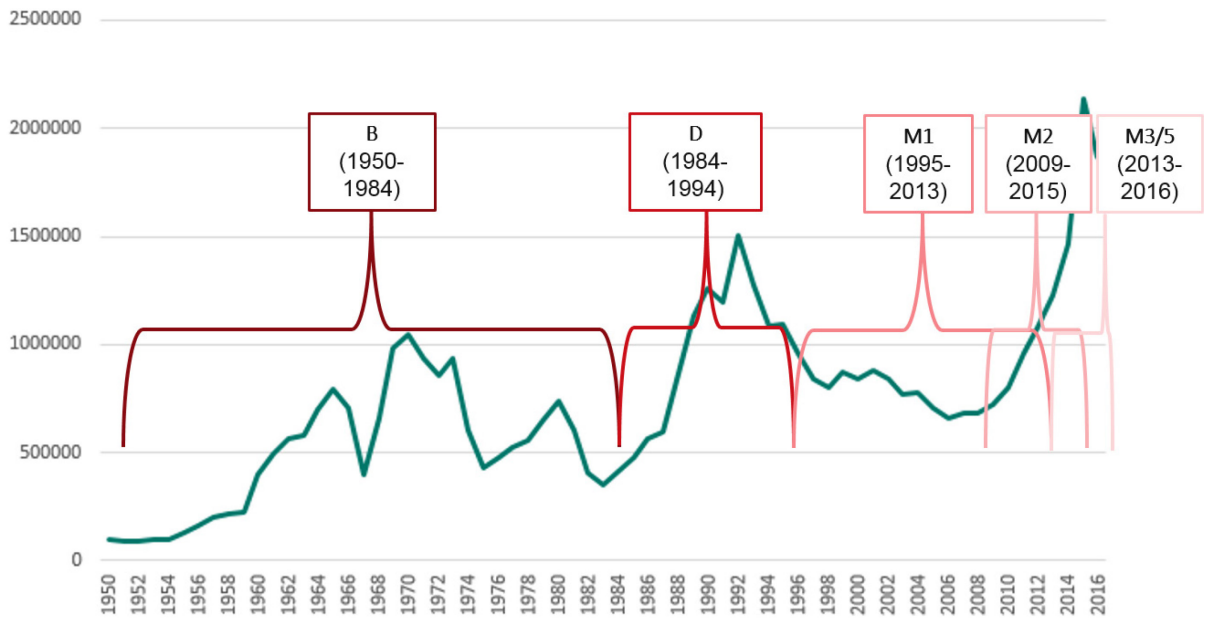

Figure 4: Gross migration to Germany and corresponding SOEP migrant sample
(DStatis, 2019)

Figure 4 displays the migrant samples added to the SOEP in response to immigration-related changes in the target population. Whereas sample B in 1984 is a general sample of migrants who had arrived in west Germany up to the start of the SOEP, sample D focuses on migrants who arrived between 1984 and 1995. Additional migrant samples in 2011 and 2013 fill in gaps in later migration. Figure 5 displays the share of respondents with a direct (personal) or indirect (parental) migration background in each survey year. As a result of the efforts of the SOEP, the graph indicates that throughout its history, respondents with a migration background have always been a substantial part of the sample population. As a result, the SOEP allows for robust analyses.

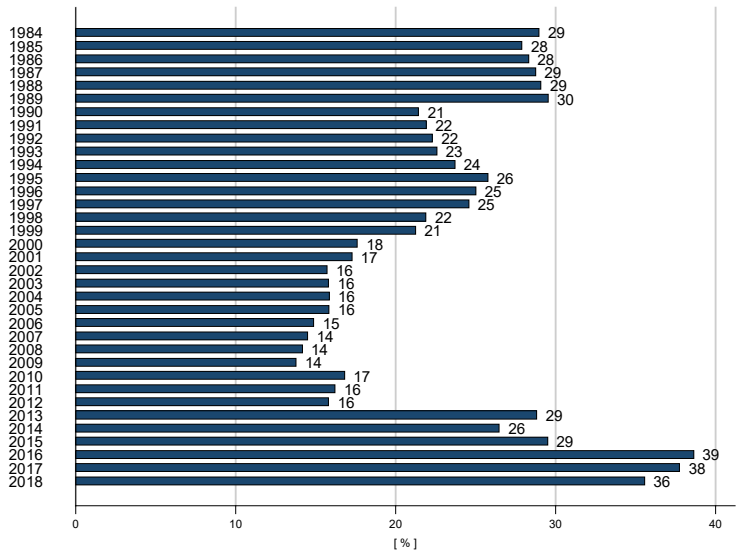

Figure 5: Share of respondents with a migration background in the SOEP since the beginning of the survey
Own calculation based on N = 142,308 individuals, SOEP v.35 (SOEP, 2019)

In order to fulfill its obligations and maintain its tradition, the SOEP needed the 2013-2016 cohort of refugees to be part of its target population. However, including them posed a major challenge for two reasons: first, their arrival was to some extent unforeseen and rapid, but they were still substantial in number. Second, the social structure of this refugee cohort differed considerably from earlier migrant cohorts in terms of the country of origin. Therefore, despite the SOEP's nature as a household panel, it was unlikely that these refugees would move into existing panel households. In order to fulfill the SOEP's mission of providing information about individuals and households in Germany, there was a need to quickly add a sample of recent refugees. As will be shown in Chapter 2, the research consortium drew significantly on the SOEP's past experiences with implementing migrant samples. However, some challenges remained, and new ones arose.

1.3 Scope

Based on the history of the SOEP and the unique inflow of refugees to Germany, this dissertation will discuss and analyze obstacles that emerge in the implementation of a survey of refugees and in the use of these data in research on integration. As I will show, such obstacles may be analytical but also methodological in nature. Therefore, by writing this thesis, I want to show that, for the purpose of research on the integration of refugees, not only the analysis of markers and means is important; the consideration of survey methodological issues is crucial as well. I argue that based on surveys the analysis of integration starts not with observing behavior but with the sampling process, and ends with the analysis and

interpretation of the data. This thesis therefore examines different aspects of a quantitative research process: 1) sampling, 2) response, 3) validation of concepts, 4) analysis and interpretation of data. The design of this thesis is cumulative, and incorporates four articles (Chapter 4–7) that deal with the four aforementioned topics:

1) The first article introduces the sampling process used in the IAB-BAMF-SOEP Survey of Refugees: a novel approach to sampling refugees in Germany. With co-authors, I propose a sequential sampling strategy to survey mobile populations such as asylum seekers and refugees. We additionally discuss obstacles encountered during the field phase, including the translation of questionnaires.
2) The second article deals with the field phase, discusses the translation of questionnaires in more depth, and examines the issue of bias in surveys due to the lack of questionnaires in respondents' mother tongues. The paper indicates that the lack of a questionnaire in respondents' mother tongue leads to increased item nonresponse. Additionally, I show that audio recordings of written questions do not help in decreasing item nonresponse.
3) The third article deals with operationalizing latent constructs, a common tool used to observe complex social realities. Taking respondents' answers on conceptions of democracy as an example, the paper discusses how different questionnaire languages and differences in political socialization hamper comparability.
4) The fourth article focuses on the analysis of economic integration and asks whether it is adequate to understand the integration of refugees mainly as an individual choice that is independent of the institutional framework in which it occurs. Analyzing the German Federal Law of Recognition *(Anerkennungsgesetz des Bundes)*, I argue that especially residence permits as well as integration classes are important to look at to understand integration trajectories. I further argue that not only labor market participation is a useful indicator of economic integration but also usage of institutions such as credential recognition.

In these four articles, which are either published, in revise and resubmit, or submitted for publication, I trace major parts of the process of implementing this survey, from data collection to analysis.

This thesis contributes to two strands of literature. First, by identifying methodological pitfalls in a survey on refugees, it contributes to the literature on cross-cultural survey methods. Second, by discussing the role of institutions in integration processes, it contributes to the literature on refugees' economic integration.

Besides adding to the research on cross-cultural survey methods and refugee integration, I want to push forward a discussion aimed at bringing together survey methodological and empirical research. Looking at the many articles that deal with the integration of migrants, it is evident that methodological considerations are often left out (Pritchard, Maehler, Pötzschke, & Ramos, 2019). Information on the employed data is scarce and it is therefore difficult to verify whether the data provide a basis for answering the research questions. Furthermore, sources of bias that emerge during the field phase or questionnaire development, are usually not addressed either (Mustillo, Lizardo, & McVeigh, 2018). Additionally, a great deal of survey methodological research fails to formulate how it can be of help in specific research questions. Many articles remain abstract, making it difficult for scholars from other fields to apply survey methodological insights. In light of these issues, I argue that there should be no differentiation between survey methodological and content-related analysis, because both fields serve the same end: creating evidence-based knowledge.

Finally, over the past four years, together with my colleagues at DIW Berlin I have contributed to the research on integration of refugees and cross-cultural survey methods in other published works that are not part of this dissertation but have the same thematic focus. For example, we investigated whether a panel app reduces unit nonresponse in a survey of refugees (currently under revise and resubmit at *Social Science Computer Review*). Other articles document sampling, weighting, and questionnaire translation in the Socio-Economic Panel (Jacobsen, 2018; Jacobsen et al., 2019; Kroh, Kühne, Jacobsen, Siegert, & Siegers, 2017) or deal with labor market access of refugees (Jacobsen, Krieger, & Legewie, forthcoming; Jacobsen, Kroh, Legewie, & Salikutluk, 2018) or describe novel field experiments that were included in the IAB-BAMF-SOEP Survey of Refugees (Jacobsen, Jaschke, et al., forthcoming). Additionally, with other colleagues from the field of integration research, I received a stipend to develop general ideas around refugee integration in Germany (Böhmer et al., 2018). My curriculum vitae in the dissertation appendix provides a comprehensive overview of all my scientific contributions.

2 The Research Question: What Challenges Arise in Sampling and Measurement when Implementing a Survey of Refugees?

This section derives the research question of this thesis. In section 2.1 I will focus on survey methodological aspects and in section 2.2 on issues regarding empirical integration research.

2.1. Challenges in Sampling and Measurement

The integration of refugees in Germany has drawn widespread attention since 2013. Questions about markers and means of integration have been heatedly discussed in both academia and the public. However, in Germany, there is no comprehensive and openly accessible register of refugees providing answers to these questions. Due to this lack, the research on refugee integration in Germany has to rely on surveys.

There are two major ways to design social surveys: as panel studies and as cross-sectional studies. While cross-sectional studies usually have a clear-cut thematic focus and are generally designed to provide timely information on few specific topics (e.g. the Eurobarometer, 2020), panel data are more powerful in allowing causal inferences (Chamberlain, 1984; Giesselmann & Windzio, 2012). Therefore, to analyze integration trajectories of migrants, the use of panel data is inevitable. Planning such surveys is by no means a trivial endeavor (an overview is provided in Harkness, Van de Vijer, & Mohler, 2003).

The most crucial point for surveys is that sampling must be random and that the sampling frame needs to contain the target population in order to ensure generalizability (Häder & Gabler, 2003; Kish, 1965). This is because the most common statistical tools and estimates rely on the central limit theorem (Ferguson, 1996), which postulates that point estimates of random samples distribute normally around the true but unknown mean of the population. Researchers exploit this assumption to carry out inferential statistics and to estimate confidence intervals and standard errors. However, even if sampling is random, further obstacles arise and have been described in the literature (see also literature review in chapter 3 on survey quality). These occur, for instance, during fieldwork and measurement (Harkness, 2003; Kalgraff Skjak & Harkness, 2003).

Until the IAB-BAMF-SOEP Survey of Refugees was launched, no panel study on refugees existed in Germany. As a result, the research consortium had only limited knowledge and experience it could draw on for the implementation of the survey. From a survey

methodological viewpoint, creating a sample of migrants in Germany is far from trivial. This is for two major reasons. First, there is only one database in Germany that incorporates all non-German citizens: the AZR, or central register of foreigners (Babka von Gostomski & Pupeter, 2008). However, due to data regulations, access to the AZR is heavily restricted and has only recently been provided to the SOEP. Second, if migrants begin the process of naturalization, they can no longer be identified, as there is no comprehensive register database on German residents. To cope with these issues, the SOEP used to apply various techniques to sample migrants such as onomastic procedures (on sample M1, (Kroh, Kühne, Goebel, & Preu, 2015)) or even respondent-driven sampling (on part of sample D, (Schupp & Wagner, 1995)). Due to the great public interest in reliable data on refugees, the Federal Office for Migration and Refugees changed its policy in 2015 and made it possible for the SOEP to sample refugees, marking the first time a third party was granted access to this unique data set for sampling purposes. However, as I will show over the course of this thesis, even though the sampling frame allowed proper random sampling, other challenges arose.

Table 2 summarizes the different sampling techniques used in the SOEP to sample migrants. As depicted, different designs and frames have been utilized to date. One set of challenges arise from the different sampling designs and sampling frames. Respondent-driven sampling (e.g., used in part of sample D) does not allow for identifying the target population, and thus design weighting is impossible. Additionally, applying ADM (*Arbeitsgemeinschaft Deutscher Marktforschungsinstitute*) networks is expensive: one cannot explicitly identify migrants but has to go through extensive screening interviews. Using the Integrated Employment Biographies, or IEBs (samples M1 and M2), leads to undercoverage since the IEBs only cover people who have either been employed or received social welfare benefits at least once (Zimmermann, Kaimer, & Oberschachtsiek, 2007). Furthermore, onomastic procedures are selective in identifying migrants and can produce false negative results (Liebau, Humpert, & Schneiderheinze, 2018). Thus, by using the AZR as a sampling frame, the research consortium expected to tackle some of the problems the SOEP had faced in the past. Moreover, as the study's target population consists of recent refugees, who could not yet have been in the process of naturalization, the AZR as a sampling frame should not be prone to undercoverage.

Nevertheless, as it turned out, the AZR was subject to an "EASY gap" (EASY is the *Erstverteilung der Asylbegehrenden*, the system for initial distribution of asylum seekers). The EASY gap refers to the fact that, although asylum seekers were registered at the border, they only appeared in the AZR at some delay.

Table 2: Migrant Samples in the SOEP

	B (1984)	D (1995)	IAB-SOEP (M1, 2013)	IAB-SOEP (M2, 2015)	IAB-BAMF-SOEP (M3-M5, 2016/2017)
Target Population	Private households with Turkish, Greek, Yugoslavian, Spanish, Italian head of household (based on citizenship) up to 1984	Migrants to West Germany between 1984 and 1995	Immigrants to Germany between 1995 and 2011 and second-generation migrants born after 1976	Immigrants to Germany between 2009 and 2013	Asylum seekers and refugees to Germany between 2013 and 2016
Sampling Frame	Register of Foreigners on County level	ADM Networks (Arbeitskreis Deutscher Markt und Sozialforschungsinstitute)	Integrated Employment Biographies (IEB) with additional onomastic procedure	Integrated Employment Biographies (IEB)	Central Register of Foreigners on Federal Level
Translation of Field Material	Turkish, Greek, Yugoslavian, Italian, Spanish	Turkish, Greek, Yugoslavian, Italian, Spanish	English, Russian, Turkish, Rumanian, Polish	English, Russian, Turkish, Rumanian, Polish	English, Urdu, Pashto, Fari/Dari, Kurmanji, Arabic

Thus, the sampling frame was updated on a regular basis. The sampling design needed to account for this issue, which had not been dealt with before. This thesis will fill this research gap (see chapter 4).

Besides sampling, the research consortium faced additional challenges in the implementation of the panel study on refugees. Two of these challenges were identified to potentially have a major impact on survey quality: *questionnaire translation*, and *selection of adequate items*. Although the SOEP has used translated questionnaires before (see again table 1), the consortium expected that these were not sufficient as the new refugees were from different countries than the previous migrants. Additionally, prior to questionnaire development, there was little knowledge on how the target population would respond to measurement instruments developed in Germany. Some concerns arose that, due to cultural differences, questions could be understood differently than assumed or that some questions would simply not make sense to refugees in their situation. Therefore, for the refugee survey, a qualitative pretest was conducted to learn about this new target population (Brücker et al., 2016). However, analyses of whether the fieldwork and instruments introduced error could only be carried out ex-post.

To give an example of the issues identified in ex-post analysis: Respondents can only answer properly if they understand the content of the questions. To ensure comprehension, questionnaires should be available in a variety of languages. As I will demonstrate in this thesis, even though the research consortium of IAB, BAMF, and SOEP provided a wide range of different languages, some respondents still could not answer in their mother tongue. As effects of a language mismatch have not been investigated to date, I will fill this research gap in the following (see chapter 5 in this thesis).

Even if all respondents understand the questions correctly, cognitive processing during the answering process (Tourangeau, Rips, & Rasinski, 2000) might differ between respondents due to socialization and thus lead to incomparability of different constructs that are measured in the questionnaire. Potential cross-cultural incomparability has only recently gained attention in the sociological literature. Taking survey items on conceptions of democracy as an example, I fill a research gap regarding the question whether value conceptions are comparable cross-culturally and cross-linguistically in the

refugee population, and in comparison to the German population (see chapter 6 in this thesis).

Based on my analysis of these issues, I identify three major methodological challenges that arise when surveying refugees:

1) Sampling when using a constantly updated sampling frame
2) Language issues during fieldwork
3) Cross-group comparability of measurement

2.2 Integration of Migrants – Changes over Time

Not only from a methodological perspective but also in terms of societal narratives and perceptions of migrant integration, the SOEP is an outstanding source of data for analyzing issues and innovations in integration research. The different attributions for migrants (e.g., direct vs indirect migration background) are a vivid example of this (for an overview see Bade, 2000; Bjanesoy, 2019; Oltmer, 2016). Although the SOEP already included a sample of migrants with its starting wave in 1984, only since the year 2008 has it provided a variable for respondents' migration background (direct, indirect, no migration background). Before that, information on migration history was derived from the country of birth or citizenship (Liebau & Tucci, 2015). This illustrates how the definition and perception of migrants has changed. Especially with the influx of guest workers in the early 1960s, being a migrant was synonymous for being foreign-born. Only after some time did this perception change, starting with the second generation of guest workers who were born in Germany but did not have German citizenship due to naturalization regulations. This made a distinction between direct and indirect migration background necessary. Figure 6 shows the composition of the population of people who are living in Germany and considered to have some sort of migration background.

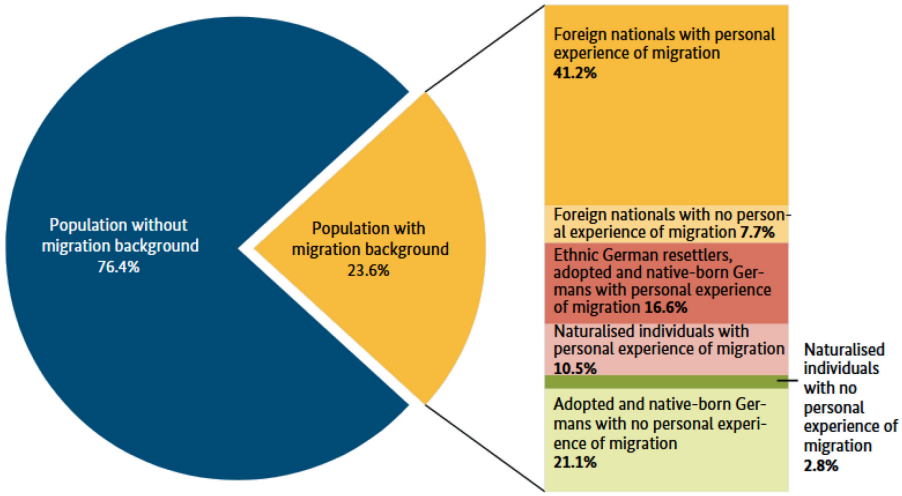

Figure 6: Share of German residents with migration history
(BAMF, 2019, p. 13)

Figures 7 additionally shows how the share of respondents with an indirect migration background in the SOEP refresher and enlargement samples has changed over time. The increase in respondents with an indirect migration background, especially between samples E, F, H, and N, which did not explicitly over-sample migrants, is an indicator of the increasing importance of this phenomenon.

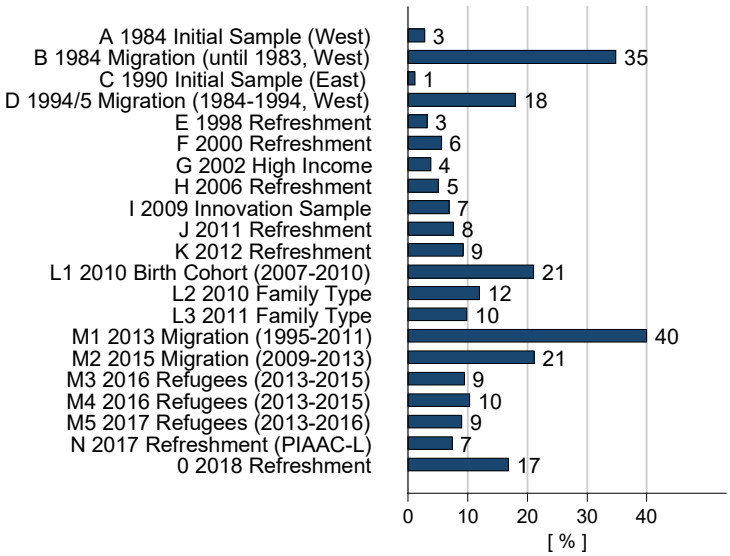

Figure 7: Share of indirect migration background in different SOEP samples in 2018
Own calculations based on N = 142,308 individuals, SOEP v.35 (SOEP, 2019)

Not only the definition of "migration background" but also the predominant narratives of migrants have changed fundamentally: from guest workers to ethnic Germans to

refugees (Seifert, 2012). It is striking that these designations are sometimes ambiguous or even appear arbitrary. For instance, displaced ethnic Germans after the second world war (*Vertriebene*) were technically not refugees, but were perceived as such (Böke, 1996), and guest workers at some point were referred to as people with a migration background, although their status did not change. These changes are also reflected in official legislation. Naturalization was progressively liberalized along with the criteria for obtaining a work visa (Goebel, 2019) and the right to obtain recognition of foreign educational credentials (Jacobsen, 2019). This variation and ambiguity in official legislation and unofficial narratives briefly illustrates how research designs and hypotheses on migrant integration trajectories have to be adapted over time to observe integration trajectories properly.

Therefore, when conducting research on the newest cohort of refugees, scholars need to question whether assumptions that were adequate in integration research in the past still apply today.

Such changes can also be observed in the SOEP fieldwork. It seems astonishing that with the initial migrant sample B and again with sample D, only a small share of questions (6 out of 68 in sample B) in the individual questionnaire were designed for the purpose of migrant-related integration research. Most field instruments were taken from the general population surveys of the SOEP (Wagner et al., 1994). This illustrates that research on migrants to that time was not concerned with questions regarding (social) integration (e.g. language, educational aspiration, culture) but rather with the impact on economic parameters. With the changing narrative of migrants and their increasing importance in the German society, however, the SOEP has increasingly found ways to implement instruments that are specifically designed to observe integration trajectories. In the new migrant samples (starting with M1), the SOEP implements extensive item batteries to understand the course of immigration, pre-migration experiences, and current attitudes towards integration.

This brings me to the conclusion that, besides methodological challenges, researchers on refugee integration face additional obstacles when implementing a study on refugees: they must formulate appropriate assumptions about integration trajectories, implement these assumptions in a questionnaire, and use them in analyses. As briefly shown, narratives and frameworks change over time and differ between different types of

migrants. Thus, assumptions and their measurement in integration research have to be constantly questioned and adapted (see chapter 6 and 7 in this thesis). One such assumption can be seen in regard to forced migration. Refugees differ substantially from other migrants in two respects: They do not migrate voluntarily, and their residence permit is not linked to their integration efforts (language proficiency, labor market participation, income) but rather to their reason for fleeing their country of origin. The analyses of integration trajectories should acknowledge these differences to ensure that analyses is not biased (see chapter 7 in this thesis). For example, due to their involuntary migration the role of institutions and with it support for integration might be more important as refugees have no time to get information on the receiving country prior to immigration.

In sum, for researchers studying refugee integration, working with the IAB-BAMF-SOEP Survey of Refugees as a supplement to the SOEP offers two important advantages: First, survey methodological challenges and solutions are tested and documented, and, second, assumptions on the integration trajectories of refugees are constantly being refined. This thesis thus seeks to answer the following research question:

What challenges arise in sampling and measurement when establishing a survey of refugees?

3 Theoretical Frameworks for and Survey Methodological Pitfalls of Measuring Integration of Refugees in Germany

This section provides a literature review on the two research fields to which this thesis contributes in order to answer the research question: these two fields are, first, research on the integration of refugees, and, second, research on survey quality. By reviewing the previous research, I will point to the research gaps, this dissertation aims to fill. The first part focuses on the migration-specific literature on integration trajectories. The second part gives an introduction to the research on survey quality, focusing on the survey lifecycle paradigm and the total survey error framework to identify possible methodological pitfalls in the research on refugee integration.

3.1 From Assimilation Theory to Ethnic Boundary-Making

This section provides an introduction to migrant-specific integration theories. I will start by introducing classical micro-sociological approaches because they are the starting point for the sociological debate on integration trajectories. I will connect these approaches to theories that have a stronger macro-sociological focus in light of the growing importance of institutional aspects in explaining integration trajectories (e.g. Ager & Strang, 2008; Kreisberg, 2019; Strang & Ager, 2010). I will conclude by discussing important recent developments in integration research, which point to the need for integration research to take a broader focus that encompasses but also looks beyond migration.

In the last few years, there has emerged a growing body of literature explicitly addressing the integration of refugees. However, in the quantitative literature, this phenomenon is quite young (Pritchard et al., 2019). Prior to this development, most research dealt with labor migrants, late expatriates (especially in Germany), or family migration (as examples see Kalter & Kogan, 2014; Krieger, 2019; Salikutluk, 2016; Shen & Kogan, 2019), while very little literature dealt specifically with the integration of refugees. As a result, most of the theoretical work on integration trajectories was developed with voluntary migration in mind. Refugees, however, usually do not leave their country of origin by choice and have very little way of planning exactly where they will eventually end up.

Additionally, for refugees, in contrast to labor migrants, residence permits are not granted based on the individual's integration efforts. This means that refugees in Germany (and in all other countries that are signatories to the Geneva Convention on the Status of Refugees) do not have to prove employment, language proficiency, or income in order to obtain a visa. That being said, one must also assume that for refugees, integration trajectories follow a different logic than for labor migrants. The following review of the integration literature should be viewed in light of these differences. I therefore conclude this review with a corresponding classification of the approaches presented.

Research on integration in the scientific literature is voluminous and hotly contested. A search for the term "integration" on Google Scholar in November 2019 produced over 6 million hits. Searching for "integration migrants" still got almost 800,000 hits, with 40,000 since 2015 alone. One observation is striking that when working through the sociological literature on integration with a focus on migration. There are two major approaches to understand and conceptualize the term "integration": On the one hand, there are micro-sociological approaches, which conceptualize "integration" as reciprocal individual behavior. These approaches are rooted in the assumption that society is the product of social behavior and that individual action is at the core of sociological analysis (Coleman, 2001; Esser, 1993; Greve, Schnabel, & Schützeichel, 2008). On the other hand, there are macro-sociological approaches, which additionally focus on institutional aspects of integration such as citizenship and rights. They conceptualize "integration" as the interplay between individuals and institutional settings. This entails that societal behavior, by definition, cannot be analyzed independently of the institutional framework (Helle, 1985; Shils, 1975).[2]

These two schools of thought are linked to two normative perspectives, each grounded in one of the two theoretical frameworks. First, focusing on integration as an individual behavior and effort puts the responsibility for integration to the newcomer. Second, focusing solely on institutions places most of the emphasis on discriminatory practices

[2] However, macro-sociological approaches should not be confused with holistic approaches, which argue that society is more than the sum of its parts (i.e., individual interaction). Macro-sociological approaches instead focus on systems and institutions that are the product of individual behavior. While the holistic approach might be interesting for theoretical debates, it is not applicable to empirical research, which focuses on the observable.

and legal regulations and thus risks impeding differences in the observation of integration trajectories that are due to differences in individual behavior.[3] In the end, research on the integration of migrants risks becoming a normative endeavor, and the development of theoretical approaches always must be viewed in light of the normative implications of migration.

In the following, I provide an overview of both approaches and how they have evolved historically within academic research (comprehensive overviews are also provided by Hans, 2016; Kalter, 2008; Nauck, 2008).

Microfactors of Integration

Under the umbrella of micro-sociology, many different integration theories have evolved over time. The classic assimilation approach was one of the first theories to explain how newcomers and a host society grow together. The notion of "segmented" or "downward" assimilation enhanced this approach by emphasizing that migrants often find themselves in lower classes of the host society. New assimilation theories took this observation and refined the classical approach by acknowledging assimilation as a reciprocal process. More complex theories have been developed recently that do not only look at assimilation processes but also describe the coming-together of migrants and host society as more complex process with various outcomes, one of which is assimilation.

The c*lassic assimilation theory*, coined by the Chicago School (an introduction is proivded by Park, 1950; Park & Burgess, 1921), was the first sociological approach that systematically explained the interplay between migrants and host society. Historically, this school of thought evolved during the wave of migration to the US in the early twentieth century, prior to the Second World War. The way migrants behaved and adapted to the new society had major impact on the theoretical developments at that time. As most immigrants came from European countries and were therefore culturally close to the host society, scholars of the Chicago School argued that migrants would

[3] These two assumptions highlight an ideological conflict that often underlies the political classification of migration inflows: Is it solely migrants who need to integrate and adapt to the host society, or do societal institutions also need to change so that migrants have access to the society?

ultimately assimilate to the host society's middle stratum (often referred to as mainstream society).

The classic assimilation theory is a unidimensional approach and understands integration as an assimilation process. It assumes that migrants at some point detach from their culture of origin and incorporate "the memories, sentiments, and attitudes of other persons or groups [in the host society], and, by sharing their experience and history, are incorporated with them in a common cultural life" (Park & Burgess, 1921, p. 735). In turn, according to this theory, all individuals would have equal access to societal participation despite having different ethnicities (Hans, 2016, p. 29). Describing migration in Israel, Eisenstadt, using the term "absorption" has developed as similar theory. In contrast to the Chicago school, however, he additionally focuses on the circumstances under which people decide to migrate. From his point of view, immigration is often the last option in people's life to solve individual problems such as unemployment or unfulfilled educational aspirations. Absorption or assimilation is then the process that takes place after immigrating to a new country (Eisenstadt, 1954).

In 1964, Gordon refined the assimilation approach (Gordon, 1964). Instead of describing assimilation unidimensionally and as a single, unified process, he argues that assimilation can be split up into seven distinct processes (Gordon, 1964, p. 69):

1) Adaption of rules and habitus (acculturation)
2) Participation in institutions (structural assimilation)
3) Interethnic (intimate) relationships
4) Identification with the host society; national consciousness
5) Political participation
6) Absence of ethnic stereotypes
7) Discrimination due to ethnicity

Gordon's major contribution to assimilation theory is the disentanglement of different assimilation processes, which make it possible to observe different facets and outcomes of assimilation. Although Gordon differentiates among these processes and thus tries to incorporate the manifold nature of assimilation, the main argument of classical assimilation theory prevails: in the end, migrants will assimilate to the mainstream society.

However, as it turned out, this assumption did not hold empirically. The classic assimilation theory was contested and consequently refined (Portes & Rumbaut, 2001; Portes & Zhou, 1993; Zhou, 1997). Scholars observed that a large share of the cohort of migrants **after** the Second World War who came from economically developing countries did not only assimilate to the US middle class but to lower classes as well. This observation was subsumed under the term "segmented" or "downward" assimilation (see also Esser, 1999; Farwick, 2011; Granato & Kalter, 2001). In contrast to the classic assimilation approach, scholars now observed that depending on the social and economic background of the migrant and the host society, assimilation has three main outcomes:

1) Assimilation to the mainstream society
2) Downward assimilation to disadvantaged strata
3) No assimilation at all

Determinants of different assimilation outcomes lay in the provision of citizenship, size and support of the ethnic peer group, experience of discrimination, economic opportunities, and individual factors such as human capital (Farwick, 2011; Hans, 2016, p. 38).

Reacting to their critics by using the term *new assimilation theory,* some scholars tried to refine the classical approach without neglecting its main argument that migrants at some point align themselves with the mainstream society (Alba, 2008; Alba & Nee, 1997, 2003; Perlmann & Waldinger, 1997; Waters & Jiménez, 2005). As an advancement on earlier assimilation theories, researchers acknowledged that assimilation is a process that both groups go through: migrants and mainstream society. Based on this assumption, they postulated three different outcomes of assimilation processes, taking into account that the mainstream society, due to migration, changes as well (Alba & Nee, 2003; Hans, 2016):

1) Boundary crossing, when migrants are no longer perceived as migrants
2) Boundary shifting, when the host society incorporates cultural characteristics of the migrants
3) Boundary blurring, when ethnic differences lose their relevance in explaining disparities between migrants and non-migrants

The new approach acknowledges that it is not only the immigrant who assimilates after migration but also the host society that changes as well. Nevertheless, the core argument seems to prevail: migrants and mainstream society converge until differences are irrelevant or nonexistent (Hans, 2010).

Despite the constant enhancement of assimilation theories, most still face criticism for understanding integration only as an assimilation process between host society and migrant. Efforts at developing more complex theories of integration that go beyond the notion of assimilation were made by Berry and Esser. I will describe their two very similar theories of social integration (Esser, 2006, 2009) and acculturation (Berry, 1997) in the following.

Integration as Utility Function

Berry conceptualizes integration by focusing on acculturative stress as the dependent variable, thus placing emphasis on the subjective well-being of migrants (Berry, 2003). A successful transition to the host society would then be defined as the absence of acculturative stress. From his point of view, there are four different strategies that can lead to the absence of acculturative stress.

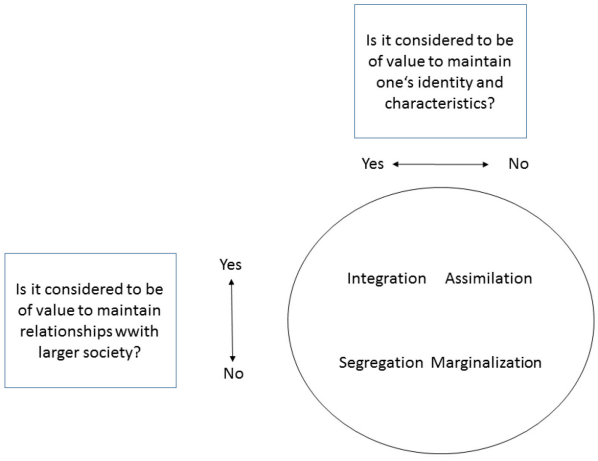

Figure 8: Berry's model of acculturative strategies
(Berry, 1997, p. 10)

According to Berry's general argument, migrants must decide how strongly they want to stay in touch with their culture of origin or adapt to the new society (see Figure 8). Different strategies can then be subsumed as acculturative strategies: integration,

marginalization, assimilation, and segregation. Integration means that migrants incorporate some habits of the new society while keeping other habits from their culture of origin. Marginalization is a strategy where both the culture of origin and the new culture are rejected. Assimilation means that habits of the culture of origin are completely disregarded and the migrant completely takes on the habits and values of the new society. Separation is the opposite, and describes a rejection of the host society while simultaneously maintaining all aspects of the culture of origin (Leibold & Kühnel, 2016). Empirically speaking, there is some indication that integration is the dominant strategy in this regard, as immigrants with ties to the host society tend to stay longer than others (Esser, 1980; Leibold, 2006). Additionally, Berry hypothesizes that immigrants who have a larger cultural distance from the host society tend to experience more acculturative stress, as the differences that need to be bridged are larger (Leibold & Kühnel, 2016).

Berry's work, especially in social psychology, has greatly influenced the work of scholars working on the integration of migrants (e.g. Green, King, & Fischer, 2019; Mägi, van Ham, Leetmaa, & Tammaru, 2018; for a systematic review see Yoon, Chang, Kim, & Clawson, 2013). Building on Berry's notion of acculturation, Ramos et al. (2016) find that acculturation is not purely a question for the newcomer, but also for the host society. They report, for example, that discriminatory experiences change acculturation strategies of migrants (Ramos, Cassidy, Reicher, & Haslam, 2016). In a similar vein, Wals and Rudolph (2018) present some evidence that pre-migratory experiences with the political system in the country of origin also influence acculturative strategies. They indicate that migrants from more affluent countries tend to stay more in touch with the culture of their host county (Wals & Rudolph, 2018). More critically, Bourhis et al. argue that they do not find empirical support for the notion of marginalization (Bourhis, Moise, Senecal, & Perreault, 1997).

Following Berry, Esser developed a similar theory on integration. Esser differentiates between system and social integration. System integration refers to the cohesion and functioning of a system as a whole and thus describes the working of markets, organizations, or the identification with and loyalty to the system (i.e. society). Regarding migrants, the notion of system integration has mostly been applied to ethnic segregation, meaning that some parts of the society (along ethnic boundaries) do not

interact with each other (Esser, 2006, p. 30). In order to comprehend the emergence of system integration, the observation of social integration is crucial.

The notion of social integration as the second aspect of integration has many commonalities to Berry's theory of acculturation.[4] Esser thus makes Berry's theory applicable for sociological purposes. Similar to Berry's notion of acculturation, social integration takes four different forms (Esser, 2006, p. 27):

1) Multiple inclusion: integration in the host society and in the migrant society
2) Segmentation: integration in the migrant society
3) Assimilation: integration in the host society
4) Marginalization: exclusion from the host and from the migrant society

As an advancement to Berry's theory, these four forms occur in four different *dimensions*:

1) Culturation: knowledge and abilities
2) Structural placement: access to the labor market, educational system, etc.
3) Social interaction: relationships
4) Emotional identification: identification with the host society

There can be a cultural, structural, social, and emotional marginality, segmentation, assimilation, or multiple inclusion (Esser, 2006). Social conditions in the country of origin (e.g., migration flows) and host society (e.g., residence regulations) and the composition of the ethnic group (e.g., social ties and human capital) are determinants of social integration, ultimately resulting in a mixture of micro-sociological and macro-sociological factors. Referring to rational choice theory, Esser argues that these three aspects shape the context and thus the opportunities of migrants (Esser, 2006, p. 68) and that migrants start investing in culturation (e.g., learning a new language) when the utility of the investment is greater than its costs. Social integration is thus a utility function to fulfill individual preferences. For instance, for migrants with low German proficiency but a desire to work, an investment in local human capital can be less valuable than investing in intra-ethnic relationships in order to get a job in the ethnic

[4] Esser, however, uses the term integration differently than Berry. For Berry, integration is a behavior, where a migrant maintains contact with the culture of both host society and country of origin. According to Esser, social integration describes the whole process and various strategies of participation in a new society.

community where local language requirements are less important (Aldrich & Waldinger, 1990). In the end, observing social integration allows observing the emergence of system integration.

Although Esser's approach on social integration is comprehensive and thus applicable to a wide array of research questions, from my point of view it is critical that he fails to describe in depth how different preference relations come about. In one of his articles, Esser argues, for instance, that the most abstract explanations for different preferences are rooted in the "principle that all people want to maintain their mental well-being and gain social appreciation" (translation J.J.) (Esser, 2006, p. 39). This very abstract definition calls for caution, as the terms mental well-being and social appreciation are not clearly defined or put into context. Furthermore, Esser argues that the means to achieve well-being and social appreciation are determined by the market economy. His discussion of these mechanisms is relatively vague rather than actually formulating hypotheses about means of integration (Esser, 2006). Additionally, the legal frameworks for migrants vary between migrant groups and over time, meaning that the means for social integration are presumably more fluid than Esser assumes.

Especially in Germany, Esser's theory has greatly influenced the research on integration of migrants (especially so his notion on social integration). In the last few years, many have found indications that his model of migrants as rational actors is capable of explaining integration trajectories. As Esser developed his model around the question of language acquisition, it is especially popular in educational sociology (e.g. Becker & Biedinger, 2006; Kalter, 2006; Miyamoto, Seuring, & Kristen, 2018; Spörlein & Kristen, 2018). However, some studies indicate that also in other areas, e.g., name giving, immigrants' behavior can be examined through the lens of rational choice theory (Gerhards & Hans, 2009).

Scholars like Bourhis et al. enhanced the approaches of Berry and Esser by placing more focus on contextual factors (Bourhis, Moise, Senécal, et al., 1997; Bourhis, Moise, Perreault, & Senécal, 1997). They are some of the first in the classic micro-sociological and social psychological integration research to attempt to implement context factors of integration in the analysis of integration trajectories (see also Tucci, 2011). They focus on political ideologies in host countries and how these shape integration trajectories of migrants. They differentiate between four different ideologies, that shape legislation on

immigration: pluralist, civic, ethnist, and assimilation. Pluralist legislation actively promote cultural diversity through state intervention (e.g., Canada and its 1988 Multiculturalism Act). Civic legislation tolerate pluralism, however the responsibility to promote it is put into the hands of the civil society (e.g., the USA or Great Britain). Ethnist legislation impedes immigrants from becoming part of the larger society (e.g., the principle of Jus Sanguinis in Germany until 2000) and assimilation ideologies accept immigrants as a new part of the society as long as they actively adapt to the culture of the host society (e.g., the German Optionspflicht [children with two passports at the age of 21 have to emit one], comprehensive citizenship tests).

Their analyses show that the political ideology that is manifest in political legislation, proves to have an impact on integration itself, beyond the behavior of the migrant. A similar approach has been used by Kogan and colleagues, who use integration policies to explain labor market outcomes (Kogan, 2016; Kogan, Kalter, Liebau, & Coher, 2011). Similarly, based on a natural experiment in Switzerland, Hainmueller and colleagues published a set of studies emphasizing the importance of the institutional framework (e.g., naturalization) for integration trajectories (Hainmueller & Hangartner, 2013; Hainmueller, Hangartner, & Pietrantuono, 2015, 2017; for a similar approach see Kreisberg, 2019).

However, compared to the vast amount of work and theoretical developments regarding micro-factors of integration, there is little research actually conceptualizing the role of institutions within the framework of assimilation and integration theories, leaving a quantitative research gap wide open (see Nauck, 2008 for a similar argument; Strang & Ager, 2010).

Macrofactors of Integration

Although scholars like Bourhis, Hainmueller, and others have incorporated some contextual factors in their empirical analyses, contemporary assimilation and integration theories have been criticized for broadly neglecting the role of institutions in their theoretical development.

One of the most prominent efforts at filling this research gap on a theoretical level is the qualitative work of Ager and Strang (Ager & Strang, 2008). Instead of focusing solely on the individual behavior of migrants, they emphasize institutional aspects and the ways

that institutions can shape and restrict the individual behavior and decisions of migrants (a similar argument is made by Pries, 2016, p. 175). Ager and Strang were also the first to explicitly construct a paradigm of integration around the specificities of refugees. The work in this dissertation is greatly influenced by these considerations, and the fourth paper in particular builds on their research.

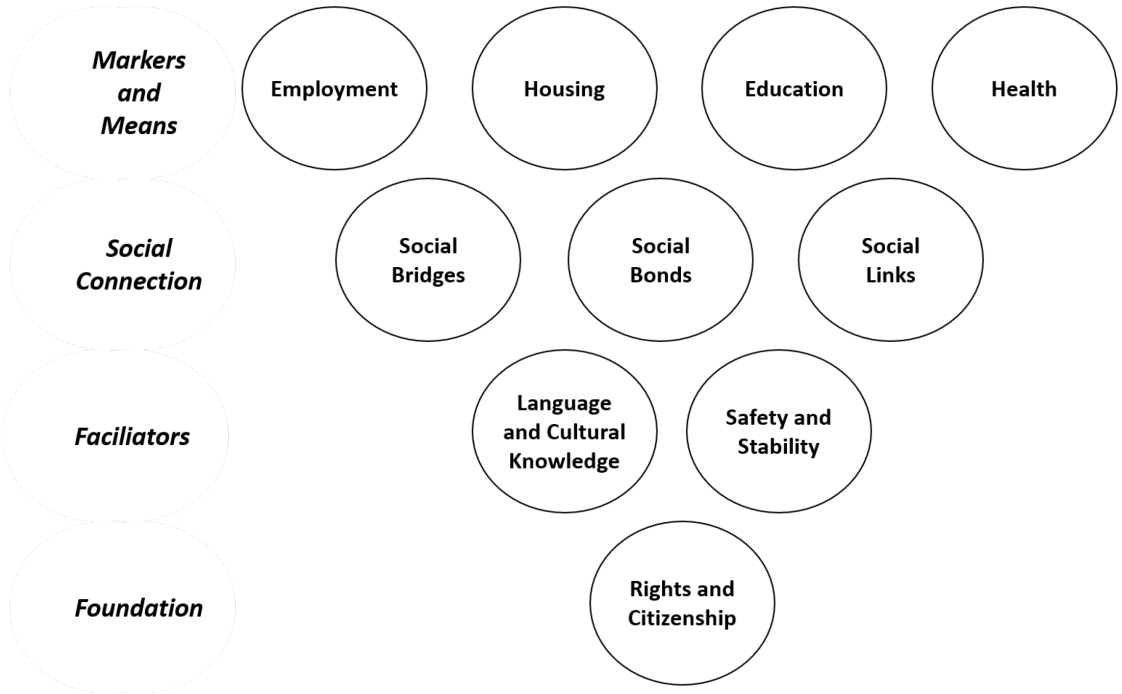

Figure 9: Indicators of Integration
(Ager & Strang, 2008, p. 170)

Using qualitative interviews with refugees in Great Britain, Ager and Strang developed a refugee-specific framework of integration (Ager & Strang, 2008). As can be seen in figure 9, they identified four layers of integration, which are expressed by means of different aspects. What stands out is that rights and citizenship are the foundation of integration, as they provide legal access to societal domains such as the labor market, welfare benefits, and political elections. Cultural knowledge and safety and stability (e.g., long-term residency) are facilitators for integration, as they reflect the cultural competencies (cultural knowledge) and planning security (safety and stability) that are necessary in order to take action. Further, social connections such as bridging ties and access to government services (social links) allow individuals to actually gain access to societal domains, which in the end is expressed through markers and means such as educational attainment, well-being, or employment.

To sum up, according to Ager and Strang, integration is not only an individual choice or an individual fate but also the outcome of an institutional setting that fosters or hampers integration. Regarding the integration of refugees, they explicitly address integration policies. In Germany, for instance, since the migration of refugees starting in 2013, integration policies have been changed constantly. Some legislation revoked rights, for instance, by suspending family reunification for people with subsidiary protection; other legislation provided additional rights, for instance, by lifting the *Wohnsitzauflage* requirement for refugees to remain living in a location determined by the government for a certain period of time. From the perspective of this integration paradigm, such legislation has both direct and indirect effects on integration trajectories.

To systematize the impacts of the legal system, da Lomba distinguished between a public and a private dimension of integration (da Lomba, 2010). While private integration refers to the refugee's actual situation (e.g., on the labor or housing market), the public dimension refers to the social and legal environment in which the refugee lives (e.g., Crul & Schneider, 2010). Da Lomba's distinction provides a basis for criticism of micro-sociological approaches: from perspective of macro-sociology, they mostly focus on the private dimension and thus insufficiently describe the whole picture of integration trajectories. This has a political and therefore normative implication: Micro-sociological approaches frame integration as an individual choice, thus putting the responsibility of integration into the hands of the migrant.

Against such convictions, some scholars argue that individual choices and desires can only be powerful when there is an infrastructure to support them (e.g. Pries, 2016; Valtonen, 2004). Da Lomba for example states that the provision of such an infrastructure for refugees would be the responsibility of welfare states (da Lomba, 2010, p. 418) – admittedly a normative assumption as well. In sum, besides the willingness to pursue integration, according to da Lomba, integration trajectories are mainly influenced and determined by the implementation of a welfare state and a migration regime.

From Integration to Inclusion: An Excursus

So far, I have presented work that directly deals with the term integration. As noted above, its definition varies within the literature: For some, integration is a synonym for assimilation, while others see integration as one of many ways of avoiding acculturative

stress, or as the result of opportunities provided by institutions. What most approaches have in common is the belief that integration is an adaptation process from one culture, society, or region to another. However, some scholars go beyond such frameworks to shift the discourse toward inclusion[5] (Böhmer et al., 2018; Sandor, 2017). The term inclusion is the preferred term in the research on disabled people, where its definition is derived from the UN Convention on the Rights of Persons with Disabilities. The term implies that not individuals but institutions need to adapt to allow access to society.

The goal of the inclusive perspective is to analyze how different societal groups are prevented from accessing social infrastructure, social networks, and access to knowledge about societal institutions, and how discriminatory institutions impede this access. Therefore, from an inclusion perspective, the existence of barriers to one group (e.g., disabled people) but not to another (e.g., people with a migration background) is an indication of discrimination. From a normative perspective, this unequal treatment is usually justified by referring to the involuntary nature of disabilities. However, scholars who work from an inclusion perspective argue that seeking refuge according to the Geneva Convention of Human rights is also not a voluntary choice (Sandor, 2017). Foroutan et al., comparing Germans born in the German Democratic Republic, Muslims, and migrants, similarly argue that the inclusion perspective makes it possible to broaden the perspective of societal participation from migrants to other marginalized groups (Foroutan, Kalter, Canan, & Zajak, 2019; Kubiak & Foroutan, 2018). From their point of view, depending on the subject of analysis, different groups can be marginalized and thus subject to exclusion (Foroutan, 2016, p. 231). For example, women as well as migrants—even though they are marginalized for different reasons and due to different mechanisms—are subject to discriminatory practices on the labor market (Altonji & Blank, 1999; Granato & Kalter, 2001; for migrant pay gap see Ingwersen & Thomsen, 2019; for gender pay gap see Wrohlich & Zucco, 2017).

The inclusion perspective is a normative theoretical framework that rejects the narrow focus on migrants as the sole subject of integration research. Due to its normative nature, this perspective has faced some critiques in the literature (e.g. Wolfsteller &

[5] In sociology, the term inclusion is also used in the context of Luhmann's theory of social systems. For Luhmann, inclusion refers to the placement of individuals in different societal positions within societal systems. However, this definition of inclusion is not used here (Esser, 1999; Luhmann, 1995).

Rädel, 2019). It has been argued, for instance, that it does not explain why migrants, as newcomers to a given society, should face the same demands as those who have been part of that society since birth (Mau, 2019, p. 219). The inclusion perspective has responded to these critiques by arguing that up to now, integration research has failed to identify why some groups are subject to exclusion and not others.

One effort to synthesize both the inclusion perspective and critiques thereof has been made by Reckwitz in his hypothesis of the "society of singularities". Reckwitz (Reckwitz, 2019) argues that modern societies focus to an increasing degree on individuality instead of group membership. Similar to the inclusion perspective, his approach therefore does not ask to which group an individual belongs but tries to observe how that individual gains access to societal domains.[6] Reckwitz' work makes it possible to bring together macro- and micro-sociological approaches to integration because he focuses on individuals and their behavior, but constantly asks how that behavior is embedded in an institutional framework (e.g., the educational system), and how the search for individuality leads to social stratification.

From Research on Integration to Ethnic Boundary Making: Changing the Perspective on Migrants

Connected to the inclusion approach is the framework of ethnic boundary making. The approach of ethnic boundary making argues that cultural differences between migrants and local populations are the result of social processes and that these, instead of integration trajectories, need to be the subject of analysis.

In response to numerous critiques of the concept of ethnic boundary making (e.g. Jenkins, 2014), Andreas Wimmer (Wimmer, 2008, 2013)—who coined and developed the term[7]—argued that classic assimilation, acculturation, and social integration theories follow what he calls "herder common sense" (Wimmer, 2009), implying that they

[6] Although this approach is somewhat similar to Ulrich Beck's notion of the "risk society" (Beck, 1986), in which he questions the continued existence of economic classes, Reckwitz has a stronger focus on culture and how culture is increasingly perceived as an individual trait and less a marker of group identity and membership.

[7] Although Wimmer coined this term, similar approaches existed previously and can be subsumed under the term "identity theory" or "group identity" (Phinney & Ong, 2007; Tajfel, 1981; Tajfel & Turner, 2004). Barth's work on boundary making (Barth, 1969) can also be seen as preliminary to Wimmer's.

perceive ethnicity and culture as naturally given, distinct, and static. Wimmer argues that only from this perspective does it appear logical to operationalize integration trajectories as the interplay between a migrant and a host society. To counter this idea, Wimmer argues that the correlation of ethnicity and culture is the product of social processes rather than being naturally given. Therefore, culture or ethnic boundaries must not be an *explanans* but need to be the *explanandum* in academic research (examples for this perspective are Baerveldt, Zijlstra, Wolf, Rossem, & Van Duijn, 2007; Gerhards & Kämpfer, 2017; Kruse & Kroneberg, 2019; Wimmer, 2009). If ignored, research produces empirical artefacts, which lead to wrong conclusions such as the dismissal of other important factors in integration besides culture (Wimmer, 2009, p. 247). Similar to the approach by Foroutan, Wimmer argues that there are other factors besides being a migrant that allow observation of (dis)integration: for example, class, gender, education and institutional discrimination. In other words, Wimmer's main argument is that patterns of social integration are not only due to ethnicity but also due to social stratification. Consequently, scholars following the approach of ethnic boundary making often look at different factors to integration. Prominent examples are institutional factors such as residence permits and education (Tabib-Calif & Lomsky-Feder, 2014), out-group behavior and discrimination (Boda, 2018), but also the need for social identity (Gerhards & Kämpfer, 2017).

Conclusion

The academic debate on integration provides a wide array of approaches to analyzing integration trajectories. The classic micro-sociological approaches were developed from unidimensional assimilation theories into complex theories of social integration and acculturation strategies that have sought to counter criticism by incorporating institutional factors and allowing variables besides ethnicity to play an important role in the analysis of integration. More recently, the question of institutional factors that hamper access to societal domains has gained increased attention and led to a debate on whether research on integration should be applied to marginalized groups as such, and not only to migrants.

The theories presented above must be viewed in the light of refugee integration research in order to make them applicable to this thesis. As mentioned, most theories were

developed with voluntary migration in mind. I see two major pitfalls that arise due to this:

1) Because refugees do not migrate by choice, and in some cases are not even able to choose their host country, it is at least questionable whether they have the same incentive to adapt to the host society. Refugees presumably are more likely to stay in touch with the culture of their country of origin to be prepared for return migration.
2) Considering the limited residence permits granted to refugees, which depend solely on their ability to prove that they have fled persecution in their country of origin and not on their efforts at integration, raises even more doubts as to whether the situation of refugees is comparable to that of voluntary migrants.

I therefore assume that assimilation theories are unlikely to work in the same way as they do for voluntary migration, as refugees in Germany cannot be sure how long they will be able to stay. Therefore, the incentive to adapt to the host society comprehensively is doubtful. Working with the inclusion perspective or ethnic boundary making, similarly, would not make right of the situation of refugees. Especially during their first time in the host country, when refugees live in shared accommodations, wait for a decision on their asylum application or start reuniting with their family, it is unlikely that they have agency, which in my view is necessary to apply the approaches on inclusion and ethnic boundary making. Both approaches work with the assumption that the object of analyses can make informed and independent choices – a perspective that does not necessarily apply to recent refugees. Moreover, working with the inclusion perspective is at risk of impeding the observation of refugee specific mechanisms of integration. It is thus not suitable to help answering the research question of this thesis.

For similar reasons, it is at least questionable whether the theories of Berry and Esser apply in the same manner as they do to voluntary migrants. First, as shown in previous studies, a large share of refugees show evidence of post-traumatic stress disorder (Brücker et al., 2019; Walther, Fuchs, Schupp, & Scheve, 2020; Walther et al., 2019). Second, there are indications that family separation is a serious issue for the mental well-being of refugees as well (Gambaro, Kreyenfeld, Schacht, & Spieß, 2018; Löbel, 2019). I therefore assume that the question of acculturative stress as the main driver of well-being of refugees is only secondary to more pressing factors, such as working through

the oftentimes traumatic experiences refugees had both prior to and in the process of fleeing their countries of origin.

Esser's theory seems to be more promising when it comes to explaining the integration trajectories of refugees, as he operates with only a few assumptions about how preferences towards social integration of migrants develop. However, as his theory builds on the assumption that migrants behave primarily as rational actors, it risks oversimplifying the situation of refugees. Because they do not migrate voluntarily, and return migration is often not an option or a matter of choice for the refugee, it would be problematic to assume that they can evaluate costs and opportunities the same way voluntary migrants can.

Therefore, in this thesis, I will rely mainly on the theory of Ager and Strang, as this theory is developed explicitly around the specificities of refugees. Additionally, it incorporates aspects of the macro and micro level in explaining integration trajectories and thus integrates aspects of both strands of sociological thinking. This theory thus offers the most compelling framework for hypotheses about the unique situation of refugees and their integration trajectories.

3.2 The Role of Survey Quality in Migrant Integration Research

So far, I have presented theoretical frameworks of integration research. However, as stated in the beginning, I argue that measuring integration starts with the data collection process. The prerequisite for measuring integration is high-quality data that allow researchers to draw conclusions on integration trajectories. Therefore, in the following, I review the literature on survey quality to identify possible pitfalls that impede unbiased analysis of integration.

This section is structured as follows: In the first part, I will provide a general overview of aspects of survey quality. In the second part, I will address aspects that need to be emphasized when implementing a survey on refugees.

What makes a good survey? It is by no means a trivial question, and would take more than one edited volume to answer (for a comprehensive overview see the editorial volume by Biemer & Lyberg, 2003). There are two major stakeholders in survey quality

(Lyberg, 2012, p. 112): the first are organizations such as survey producers[8] and their professional associations[9] and the fieldwork institutes that carry out data collection (Morganstein & Marker, 1997). Each of these institutions and organizations has their own code of conduct for producing high-quality survey data, but most of these codes refer more to research ethics (Weichbold, 2009, p. 558) in the specific field rather than defining clear-cut benchmarks of survey quality in general. The second stakeholders, who stand partly in contrast to these organizations, are data users. They are usually not involved in the process of collecting the data. Therefore, most of the aforementioned guidelines produced by the organizations lack a user perspective. Referring to Juran and Gryna (1993), Biemer and Lyberg (2003) therefore argue that good data are not only accurate data (i.e., adhering to statistical quality), but also data that succeed in satisfying the needs of researchers. Moreover, high survey quality would be a synonym for "fitness for use" (Biemer & Lyberg, 2003, p. 13; Juran & Gryna, 1993; Tayi & Ballou, 1998):

> "Quality can be defined simply as 'fitness for use'. In the context of a survey, this translates to a requirement for survey data to be as *accurate* as necessary to achieve their intended purposes, be available at the time is needed (*timely),* and be *accessible* to those for whom the survey was conducted. Accuracy, timeliness, and accessibility, then, are three *dimensions* of survey quality" (Biemer & Lyberg, 2003, p. 13).

This perspective emphasizes that survey methodology is not only guided by concerns of statistical quality, but that it also needs a user perspective. Its aim is to achieve balance between the user burden and utility, and the costs of a high-quality survey. However, the needs of researchers are manifold and depend on various aspects such as research questions, purpose of data usage, and skills. Biemer and Lyberg's definition is therefore more an abstract guideline to remind survey producers that data should be easy to access and use, rather than a precise definition of high survey quality (Biemer, 2010).

[8] Prominent examples are the German Socio-Economic Panel (SOEP), the Panel Study on Income Dynamics (PSID) in the United States, and the United Kingdom Household Longitudinal Study (UKHLS).

[9] These include the German Arbeitsgemeinschaft Deutscher Marktforschungsinstitute (ADM), the American Association for Public Opinion Research (AAPOR), and the European Survey Research Association (ESRA). The guidelines of AAPOR have become established as the most influential guidelines for survey quality and can be seen as a benchmark for survey producers.

The Survey Lifecycle

A comprehensive approach to understanding what components make a good survey is the concept of the "survey lifecycle". It describes the various aspects that must be taken into account when implementing a survey: from study management over questionnaire design to statistical analysis.

Figure 10 shows the different steps in the survey lifecycle. The survey lifecycle perspective assumes that all steps are overlapping to some extent. Interestingly, however, in the middle of the cycle lie "survey quality" and "ethical considerations", hinting at the fact that there must always be an equilibrium between what can be applied and what should be applied. For instance, for a survey of refugees, it is important to avoid re-traumatization (ethical considerations) without diminishing survey quality by discarding important instruments (survey quality).

The dissertation at hand does not deal with all aspects of the survey lifecycle but focuses on those aspects that come to the fore in a survey on refugees because techniques or instruments have to be adapted to this new target population. Therefore, I will focus on sampling design, adaptation of survey instruments, and analysis. Most other phases of the survey lifecycle, as I assume, are standardized and do not have to be adjusted or altered drastically in a survey of refugees (e.g., dissemination, harmonization, paradata, study management, contracts). Arguably, interviewer selection could be seen as crucial as well. Unfortunately, for the IAB-BAMF-SOEP Survey of Refugees, interviewer recruitment was the responsibility of an external field institute and interviewer information was only scarcely available impeding in-depth analyses (only information on age, gender, and migration background is available).

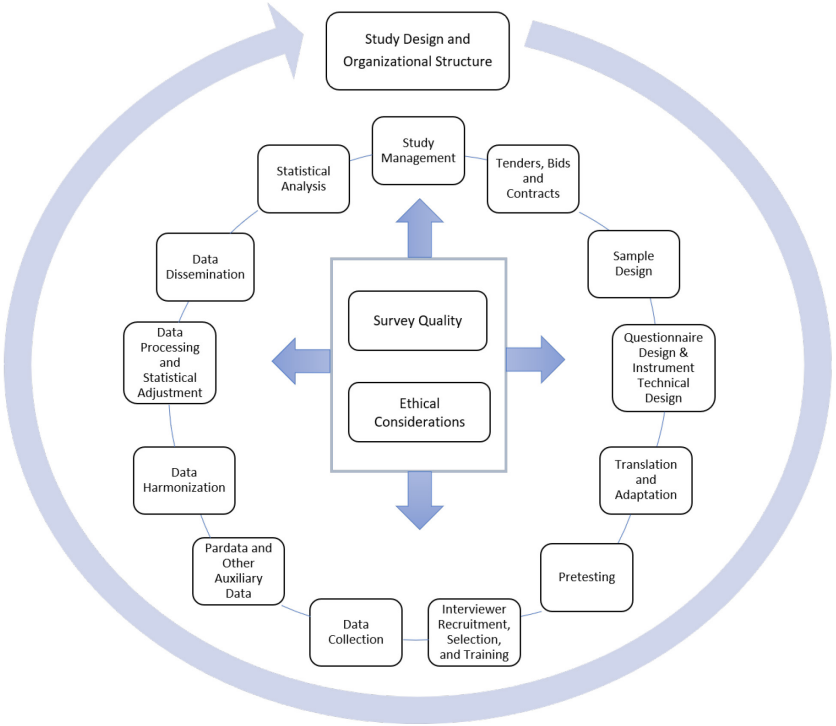

Figure 10: The Survey Lifecycle
(Hansen et al., 2016)

As mentioned, the survey lifecycle perspective does not focus on the survey methodology in particular. Therefore, similar to the notion of "fitness for use", it serves as a broad guideline and overview of the different aspects of survey implementation but does not necessarily provide practical tools to actually estimate and assess survey quality.

Total Survey Error

The aforementioned gap is filled by the total survey error perspective, a related approach that establishes clear benchmarks for assessing survey quality and is often applied in the evaluation of survey quality. To date, the total survey error perspective is one of the most influential tools for estimating survey quality. The total survey error (TSE) paradigm is a well-established concept for assessing survey quality. It was first proposed as a concept by Deming (Deming, 1944) and was subsequently applied in research on survey design and survey methods (Alwin, 2007; Groves, 1989; Hansen, Hurwitz, & Madow, 1953; Kish, 1965). According to Hansen et al. (2016, p. 750), "TSE defines [survey] quality as the estimation and reduction of the mean square error (MSE) [...] which is the sum of random errors (variance) and squared systematic errors (bias)." The mean square error thus denotes a function of:

$$MSE = \sum var^2 + \sum bias^2$$

However, the variance as well as the bias consist of several measures. I will introduce each one further below. While variance occurs at random and thus only reduces the effective sample size of surveys, bias leads to wrong estimations by definition (Smith, 2011). All sources of error, which I will mention in the following section, can be differentiated in bias and variance.

TSE usually is divided into four different sources of quantifiable error: sampling error, coverage error, nonresponse error, and measurement error. Additionally, TSE includes specification (referring to validity) and processing error, measures which are not quantifiable but relate to theoretical and technological aspects of survey production (see figure 11).

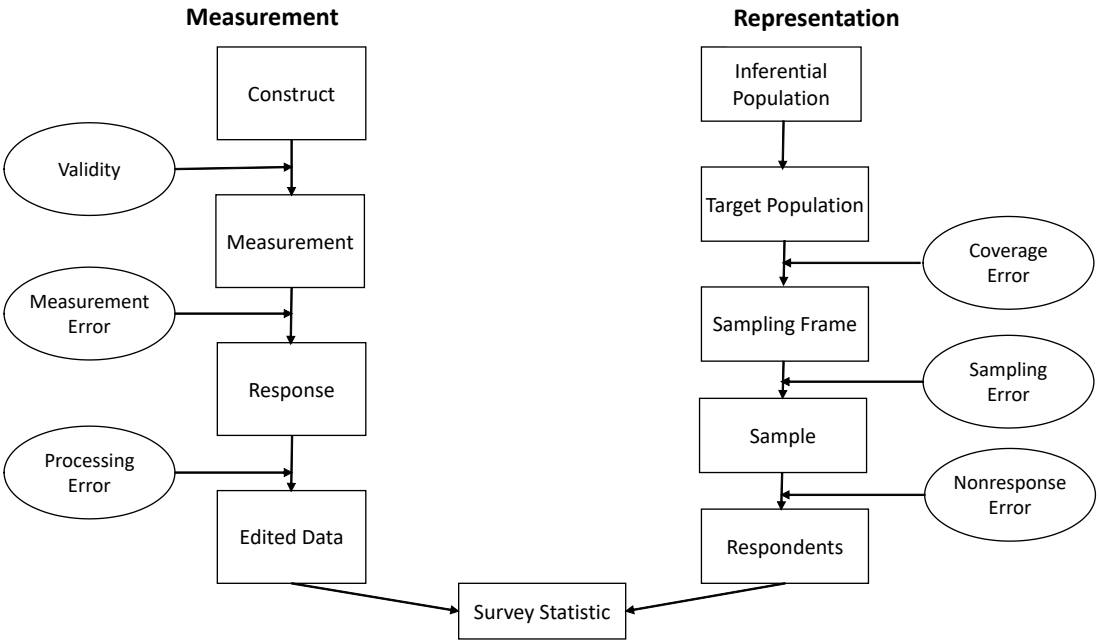

Figure 11: Aspects of Total Survey Error
(Groves, Fowler, Couper, Singer, & Tourangeau, 2004, p.49)

A comprehensive overview of the TSE approach is also provided by Bautista (Bautista, 2012).

Sampling Error

The literature on sampling errors differentiates between sampling variance and sampling bias. Sampling variance poses a challenge because it diminishes the effective sampling

size of a survey. Suppose we draw a random sample (every member of the target population has a known and non-zero probability of being sampled): due to randomization, estimates of the sample will deviate from the target population and from other random samples, which were drawn under the exact same circumstances. Thus, sampling design by definition introduces variance into survey estimates (Kish, 1965).

Besides sampling variance, a second source of variance can emerge due to sampling: a clustering effect. Similar to sampling variance, clustering diminishes the effective sample size. Cluster effects arise from the sampling strategy: Most random samples are two-stage samples. This means that before drawing cases from the actual target population, they are clustered by e.g. regions (referred to as primary sampling units, or PSU). In a first step, such regional clusters are sampled and only in a second step are the actual cases drawn from the clusters (referred to as secondary sampling units, or SSU). Because cases within clusters are assumed homogenous in nature, this diminishes the power of the sample.

Together, the design effect in the case of a sample mean *(deff*, i.e., sampling variance) resulting from the sampling design and the clustering is a function of:

$$deff = \frac{\left(\sum_{i=1}^{n} \pi_i^{-1}\right)^2}{\sum_{i=1}^{n} \left(\pi_i^{-1}\right)^1} * \left(1 + \left(n_j - 1\right)\rho\right)$$

with

$$\rho = \frac{\sigma_{PSU}^2}{\sigma_{PSU}^2 + \sigma_{SSU}^2}$$

where n = total sampled individuals (i), and n_j = average sample size in cluster *j*, σ^2= standard deviation from mean of an estimator, and π = sampling probability.

Accounting for the sampling design with proper techniques is therefore strongly recommended when analyzing survey data (e.g., clustering of standard errors, (Abadie, Athey, Imbens, & Wooldridge, 2017); random slopes and intercepts, see (Kreft & de Leeuw, 1998); or a random group concept, see (Wolter, 2007, p. 21)). However, even though sample variance can decrease the effective sample size and increase standard errors, it does not lead to wrong estimates.

Sample bias, however, does lead to wrong estimates. Usually, when drawing a random sample, every member of the target population has a known and non-zero probability of being sampled. If this is not the case, sampling bias will occur due to non-coverage of members in the target population (see also review on frame error further below). Suppose researchers draw a random sample of asylum seekers and refugees without employing a sampling frame that comprises all entities of the target population. In such a survey, it is likely that people who immigrated illegally and only filed a claim asylum after some period of time are not part of the sampling frame. As a result, some members of the target population have a sampling probability of zero and thus cannot be sampled. This results in an under-coverage of that population, and thus we cannot account for their characteristics in estimations.

Additionally, sample bias occurs if we cannot estimate design weights and thus correct for the sampling process. This problem emerges if the target population or the sampling probability is unknown. If we cannot produce design weights, the data cannot be corrected for the sampling design (e.g., disproportional sampling). Hence, estimates are biased as the sample and its characteristics do not mirror the target population. Unfortunately, if the target population is unknown, researchers cannot identify the sampling bias, and also do not know how serious the bias is. Sampling bias is therefore problematic in two ways: first, by definition, the sampling bias will lead to wrong estimates. Second, it is impossible to identify the bias from the sample itself. Researchers cannot assess whether a bias leads to small deviations that can simply be ignored, or whether the sample is biased on a larger scale.

According to Bethlehem, the sampling bias of a mean denotes a function of

$$B\overline{y_s} = \frac{R(\rho, Y)S(\rho)S(Y)}{\bar{\rho}}$$

With $\bar{\rho}$ being the average response probability (i.e. propensity), $R(\rho, Y)$ being the correlation coefficient between variable of interest and response behavior. $S(\rho)$ is the standard deviation of the response probability and $S(Y)$ of the variable of interest respectively (Bethlehem, 2002, 2010).

In sum, there are two pitfalls regarding sampling error: sampling variance and sampling bias. While the former decreases the effective sampling size, making estimates less exact (by increasing standard errors), the latter will result in wrong estimates.

Measurement Error

Measurement error refers to the fact that estimates of the sample do not reflect the true values for the respondents (Groves & Lyberg, 2010). In order to identify and quantify measurement error, a benchmark test is necessary. This implies that researchers know the true value in the target population. This assumption is quite unlikely to fulfill (otherwise, surveys would be unnecessary). However, in order to get an idea of potential bias, the literature suggests the test-retest design (Biemer, 2010). This design estimates the reliability ratio by drawing two random samples under the exact same circumstances and comparing two estimates of interest. The reliability ratio for a sample estimator (y) of respondent i denotes a function of

$$R = \frac{\sum_{i=1}^{n}(y_{1i} - y_{2i})^2}{s_1^2 + s_2^2},$$

with

$$s_1^2 = \sum(y_{1i} - \overline{y}_1)^2/(n-1),$$

whereas the same can be estimated for s_2^2 with y_{2i}. Moreover, $\overline{y}_1$ is the mean of y_{1i} (Biemer, 2010). Assuming that due to random sampling confounding factors are not an issue, if measurement error is absent, the reliability ratio should be close to one. Applying the test-retest design is thus capable of estimating whether an estimate is reliable and thus likely to reflect a true value.

Generally speaking, measurement error is defined as the difference between a true value and the estimated value for a respondent. The following four major sources of measurement error, based on TSE, are subject to discussion in the literature:

1) Questionnaire design
   The design of the questionnaire changes the way respondents answer. Question order effects (McFarland, 1981; Stark et al., 2018), respondent burden (Eckman & Kreuter, 2018; Krosnick, 1991), and social desirability (Krumpal, 2013) are three

prominent examples. A question order effect occurs if previous questions influence how respondents answer subsequent questions. Respondent burden refers to the amount of work required of a respondent to answer all questions. Frequently, respondents answer the last questions in long surveys or anticipated filter questions wrong or without actually thinking about the meaning to decrease the burden. Socially desirable response behavior describes a behavior whereby respondents, when answering sensitive questions, adjust their answers to perceived "correct" or socially acceptable answers.

2) Mode of data collection

   The mode of data collection can influence the measurement. For example, in a face-to-face interview, some respondents might not want to reveal the true answer to some questions and thus answer differently than in a situation where the interviewer is absent and respondents complete the questionnaire alone (Jann, Jerke, & Krumpal, 2011). In the literature, this phenomenon is called "mode effects" (de Leeuw, 1993; Lugtig, Lensvel-Mulders, Frerichs, & Greven, 2011; Martin & Lynn, 2011). Such effects can occur when the survey is presented differently in different modes, thus triggering diverging response behavior. The difference between web-based and non-web based modes is a prominent example of this, as the presentation of questions differs substantially (Couper & Peterson, 2016; de Bruijne & Wijnant, 2013).

3) Interviewer characteristics

   Interviewer effects regularly occur in face-to-face and telephone interviews. Such an effect comes about if characteristics of the interviewer affect the respondent's behavior. This may be due to the respondent wanting to satisfy the interviewer or because the respondent wants to present him/herself in a good light (Kühne, 2018; West & Blom, 2017). Additionally, interviewer effects are likely to occur with sensitive requests such as record linkage (Sakshaug, Tutz, & Kreuter, 2013) or requests to participate in field experiments (Legewie et al., 2019).

4) Respondent characteristics

   Some respondents might not be able to give answers in the way survey producers expect them to. This may occur due to illiteracy, diminished cognitive abilities, or physical constraints. Thus, if the questionnaire does not meet the needs of all respondents, such characteristics lead to faulty answers, because questions are

not understood properly or it is impossible for the respondent to provide a true answer (Schwarz, Park, Knäuper, & Sudman, 1999; Yan & Tounrangeau, 2008). Tourangeau and colleagues differentiate between four steps in a survey answering process: comprehension, retrieval, judgement, and response (Tourangeau et al., 2000, p. 8). Comprehension describes the respondent's processing of the question and making sense of its meaning. Retrieval describes the process of recalling relevant information to answer the question. Judgement describes the process of evaluating recalled information based on its relevance for the question. Response is the act of actually giving an answer to the question. In each of these four steps, errors can occur for different reasons. Questions can be understood incorrectly (comprehension), memories might be blurred because they were long ago or are perceived as insignificant (retrieval), respondents might think that their knowledge is not relevant for the question although it is (judgement), and the response might not fit the survey design (response).

In sum, measurement error can result from questionnaire design, mode effects, interviewer effects, and respondents' characteristics.

Frame Error

Frame error describes the discrepancy between the population researchers want to study with a sample and the sampling frame they use to draw that sample (Pennell, Harkness, Levenstein, & Quaglia, 2010, p. 277). Scholars differentiate between over-coverage (the sampling frame comprises more people than the target population) and under-coverage (the sampling frame does not comprise all the people in the target population)(Delnevo, Gundersen, & Hagman, 2008; Harter, Eckman, English, & O'Muircheartaigh, 2010; Jagers, 1986). Thus, estimating the frame error requires the mean estimate of a characteristic of the non-covered or over-covered population ($\bar{Y}_{nc}$) and the relative size (i.e. rate) of the non- or overcovered population ($\hat{Y}_{NC}$). Additionally, we need the sample mean ($\bar{Y}_c$) of the corresponding characteristic (Biemer, 2010). The frame error thus denotes a function of:

$$\hat{B}_{NC} = \hat{Y}_{NC}(\bar{Y}_c - \bar{Y}_{nc}).$$

As will be shown in Chapter 4, the tremendous under-coverage of the target population posed a serious challenge to sampling in the IAB-BAMF-SOEP Survey of Refugees. A solution for this problem is introduced in this thesis.

Nonresponse Error

Generally, the literature distinguishes between item and unit nonresponse. Unit nonresponse refers to individuals who are reluctant to participate in a survey in general, whereas item nonresponse refers only to certain questions. Unit-nonresponse often occurs when people move frequently or are busy (due to work or vocational training) and are thus hard to reach or do not have time for a survey (Groves, 2006; Groves & Couper, 2012). Item nonresponse emerges when people do not want to reveal the truth due to sensitivity or social desirability or want to finish the interview quickly (Loosveldt, Pickery, & Billiet, 2002; Tourangeau & Yan, 2007; Yan & Curtin, 2010).

In general, and similar to frame error, nonresponse error denotes a function of

$$\bar{y}_r - \bar{y}_n = (\frac{m}{n})[\bar{y}_r - \bar{y}_m] ,$$

where $\bar{y}_n$ is the full sample mean, $\bar{y}_r$ is the respondent mean, $\bar{y}_m$ is the mean for missing cases and m/n is the response rate (Couper & de Leeuw, 2003, p. 166). This function is applicable to unit as well as to item nonresponse.

Regarding cross-cultural surveys and causes of item nonresponse, the literature is scarce (Couper & de Leeuw, 2003, p. 170). The existing literature however suggests that item nonresponse is not uniformly distributed across cultures and countries, and that context effects such as socialization about privacy and values in general (for an overview see Couper & de Leeuw, 2003; Piekut, 2019) or economic conditions (Blom, Jäckle, & Lynn, 2010) have a strong impact. Thus, handling missing data in cross-national surveys has to take such contexts effects into account (for an application on labor income see e.g. Frick & Grabka, 2010).

Specification and Processing Error

So far, I have emphasized sampling, coverage, nonresponse, and measurement error. This is in line with most of the literature using the TSE framework as those errors can be quantified. However, there are two more sources of error that are frequently ignored and underestimated in their effect on survey quality: specification and processing error. Specification error occurs when there is no link between theoretical assumptions, item development, or mathematical model specifications. Thus, specification error describes

the event that the operationalization fails to reflect the research questions (Biemer, 2010; Heckman, 1979).

Processing error usually occurs at the end of data collection and is due to malfunctioning of, e.g., CAPI software, mistakes during use file preparation such as imputation and generation of variables (e.g., common classifications such as ISCED or CASMIN) (Biemer & Lyberg, 2003). Such errors can be minimized by reviewing code, discussing theoretical concepts in-depth and in-group, or by consulting experts in the specific field of research.

Methodological Obstacles in Cross-Cultural, Cross-Linguistic, and Cross-Regional Research

So far, I have presented general guidelines for high survey quality. However, when dealing with a survey of recent immigrants, additional challenges to survey quality have to be taken into account. In the following, I will highlight aspects of survey quality that need an extra focus in a survey of recent refugees.

Translation

When conducting survey research on recent immigrants, questionnaires need to be translated. However, this bears potential bias because translation is not always straightforward, as many languages (e.g., Arabic) consist of various dialects (Behr, Brzoska, & Schoua-Glusberg, 2018). If translations do not incorporate the same meaning across languages or cannot be entirely understood by the target population, this will introduce measurement error. Different measures to assure comparability between languages have been proposed, such as expert reviews and cognitive interviewing (Goerman, Meyers, & Trejo, 2018). The most prominent such measures are the cross cultural survey guidelines (CCSG) developed by Mohler and colleagues, which define a catalog of steps to ensure unbiased translations (Mohler, Dorer, de Jong, & Hu, 2016). These steps are summarized within the TRAPD model: **t**ranslation, **r**eview, **a**djudication, **p**retesting, and **d**ocumentation. According to Mohler and colleagues, for best practice, two translators working separately should produce two versions of the same questionnaire. Both versions should then be reviewed and a reconciled version is produced. The adjudication is then an iterative process between pretesting and enhancement of the translation, which eventually leads to a final version (Mohler et al., 2016, p. 235). All steps, if necessary should be repeated and documented.

Although thorough guidelines for survey translation exist, little research has been done on the absence of necessary languages during fieldwork. Thus, to date, there is a research gap on the question of how respondents behave when a questionnaire is not available in their mother tongue.

Measurement Invariance

If translation is insufficient—if it changes the meanings of questions—the measured constructs might not be comparable across groups. The lack of comparability can be tested by means of measurement invariance testing (also referred to in the literature as measurement equivalence) (Billiet, 2003). Generally speaking, researchers differentiate among three factors that lead to measurement variance: construct bias (e.g., different definitions of the construct across cultures), method bias (e.g., incomparability of samples across groups), and item bias (e.g., poor translation) (Van de Vijer, 2003, p. 146).

Applying this to a survey of refugees, especially the first and third factors seem crucial. Method bias seems less a problem in this case, as all refugees are ideally part of the same sampling frame in a national survey on refugees. Construct bias (in addition to item bias due to translation, which has been described in the previous paragraph) is crucial, as respondents in a survey on refugees come from all over the world and therefore from different political, societal, and economical contexts. A prominent example in migration research is mental health: standardized indicators used to measure mental well-being are often not comparable cross-culturally (e.g. Comanaru & d'Ardenne, 2018). A second example are attitudes. Because attitudes are, in very broad sense, the result of exposure to and experience of culture, what values encompass and refer to is understood differently in different cultures (Cho, 2014; Finkel & Smith, 2011). Huntington, for instance, argues that western societies tend to hold individualistic rather than the collectivist values predominant in Middle Eastern and Asian societies (Huntington, 1993). If this is true, values would be nearly incomparable across different cultures as they are rooted in different historic contexts (e.g. Ariely & Davidov, 2011). However, there has been no thorough analysis of value consensus in immigration societies to date, leaving an important gap in the research.

Sensitivity and Social Desirability

Questionnaire design usually takes into account that certain questions can be culturally sensitive and may therefore trigger socially desirable response behavior or item nonresponse (Krumpal, 2013; Tourangeau & Yan, 2007). In Germany, for example, reporting income, drug abuse, or voting behavior is considered sensitive (Ong & Weiss, 2000). However, in order to deal with sensitivity, scholars need to be able to identify which questions are subject to sensitivity. Research on this matter indicates that sensitivity is culturally diverse and has various causes (Schwarz, 2003). While for immigrants with a temporary residence permit, questions related to legal status might be sensitive, for religious minorities, questions on religious practices are as well. Research on recent immigrants needs to account for this in order to make sure that answers are either not subject to socially desirable response behavior or in order to apply instruments to control for social desirability bias in analyses.

There is some literature dealing with causes of sensitivity and social desirability in research on immigration. Especially questions on illegal behavior (e.g., drug abuse) have been identified to trigger socially desirable response behavior (Pennell, Hibben, Lyberg, Mohler, & Worku, 2017). Additionally, a review by Schwarz indicates that East Asian cultures tend to show a culture of fitting in and thus are more inclined to answer in a socially desirable way compared to Western (European) cultures (Schwarz, 2003). Additionally, Johnson and Van de Vijer show that the level of social desirability is negatively correlated with affluence at the national level (Johnson & Van de Vijer, 2003, p. 203). This is an indication that respondents who come from authoritarian countries are more likely to answer in a socially desirable way, which must be kept in mind when analyzing data from a survey on refugees.

3.3 Identification of Research Gaps

Based on the literature reviews on integration and survey quality I identify the following four research gaps:

1) Identifying an adequate sampling technique for refugees in times of high immigration. As discussed in the literature review, drawing random samples in which all members of the target population have a known and non-zero probability is crucial in order to avoid sampling bias. To date, such a random sample of refugees has not been drawn in Germany and thus little is known

about adequate sampling frames and specific sampling techniques. The first article in my dissertation closes this gap.

2) Language-sensitive identification of item nonresponse error in research on refugees.

   The literature review showed that there is a growing body of research on language issues in the academic literature. However, most of this literature addresses the translation process. Little to no research has been carried out on the question of how interviewees behave when questionnaires are not available in their mother tongue. This second articles aims at shedding some light on this problem.

3) Culturally sensitive assessment of measurement invariance. As presented above, in social surveys it is crucial to show that latent constructs are comparable across groups. Especially in research on well-being, this desideratum has given rise to a growing literature on comparability. In sociology, however, only a few studies to date have tested latent constructs for comparability. The third paper fills this research gap by testing value conceptions among refugees and Germans for comparability.

4) The role of institutions in refugee integration. As the literature discussion above has shown, the role of institutions in integration research has only recently become a topic of investigation. In Germany, the different forms of protection for refugees are an important institutional factor to look at when observing integration trajectories of refugees, but to date these have not been the subject of analysis. I fill this research gap by analyzing how institutions (residence permits and integration classes) affect investments in refugees' future labor market access.

References

Abadie, A., Athey, S., Imbens, G. W., & Wooldridge, J. (2017). When Schould You Adjust Standard Errors for Clustering? *NBER Working Paper*(24003). doi:10.3386/w24003

Ager, A., & Strang, A. (2008). Understanding Integration: A Conceptual Framework. *Journal of Refugee Studies, 21*(2), 166-191. doi:10.1093/jrs/fen016

Alba, R. (2008). Why we Still Need a Theory of Mainstream Assimilation. In F. Kalter (Ed.), *Migration und Integration* (pp. 37-56). Wiesbaden: VS Verlag für Sozialwissenschaften.

Alba, R., & Nee, V. (1997). Rethinking Assimilation Theory for a New Era of Immigration. *International Migration Review, 31*(4), 826-874.

Alba, R., & Nee, V. (2003). *Remaking the American Mainstream. Assimilation and Contemporary Immigration*. Cambridge: Harvard University Press.

Aldrich, H. E., & Waldinger, R. (1990). Ethnicity and Entrepreneurship. *Annual Review of Sociology*(16), 11-135. doi:10.1146/annurev.so.16.080190.000551

Altonji, J. G., & Blank, R. M. (1999). Race and Gender in the Labor Market. In O. Ashenfelter & D. Card (Eds.), *Handbook of Labor Economics* (pp. 3143-3259). Amsterdam: Elsevier.

Alwin, D. F. (2007). *Margins of Error: A Study of Reliability in Survey Measurement*. New York: Wiley & Sons.

Ariely, G., & Davidov, E. (2011). Can we Rate Public Support for Democracy in a Comparable Way? Cross-National Equivalence of Democratic Attitudes in the World Value Survey. *Social Indicators Research, 104*(2), 271-286. doi:10.1007/s11205-010-9693-5

Babka von Gostomski, C., & Pupeter, M. (2008). Zufallsbefragung von Ausländern auf Basisdes Ausländerzentralregisters: Erfahrungen bei der Repräsentativbefragung "Ausgewählte Migrantengruppen in Deutschland 2006/2007" (RAM). *Methoden, Daten, Analysen (mda), 2*(2), 149-177.

Bade, K. J. (2000). *Europa in Bewegung. Migration vom späten 18. Jahrhundert bis zur Gegenwart*. München: C.H. Beck.

Baerveldt, C., Zijlstra, B., Wolf, M. d., Rossem, R. V., & Van Duijn, M. A. J. (2007). Ethnic Boundaries in High School Students' Networks in Flanders and the Netherlands. *International Sociology, 22*(6), 701-720. doi:10.1177/0268580907082248

BAMF. (2019). *2016/2017 Migration Report: Key Results*.

Barth, F. (1969). *Ethnic Groups and Boundaries. The Social Organization of Culture Differences*. Boston: Little, Brown and Company.

Bautista, R. (2012). An Overlooked Approach in Survey Resarch: Total Survey Error. In L. Gideon (Ed.), *Handbook of Survey Methodology for the Social Sciences* (pp. 37-49). New York: Springer-Verlag New York.

Beck, U. (1986). *Risikogesellschaft. Auf dem Weg in eine andere Moderne*. Berlin: Suhrkamp.

Becker, B., & Biedinger, N. (2006). Ethnische Bildungsungleichheit zu Schulbeginn. *Kölner Zeitschrift für Soziologie und Sozialpsychologie, 58*(4), 660-684.

Behr, D., Brzoska, P., & Schoua-Glusberg, A. (2018). Linguistic and Cultural Aspects in Migrant Surveys. Introduction and Overview. In D. Behr (Ed.), *Surveying the Migrant Population: Consideration of Linguistic and Cultural Issues* (pp. 5-12). Köln: GESIS.

Berry, J. W. (1997). Immigration, Acculturation, and Adaption. *Applied Psychology: An International Review, 46*(1), 5-68.

Berry, J. W. (2003). Conceptual Approaches to Acculturation. In K. M. Chun, O. B. Organista, & G. Martin (Eds.), *Aculturation: Advances in Theory, Measurement, and applied Research* (pp. 17-37). Washington: American Psychologcail Association.

Bethlehem, J. (2002). Weighting Nonresponse Adjustments Based on Auxiliary Information. In R. M. Groves, D. A. Dillman, J. L. Eltinge, & R. J. A. Little (Eds.), *Survey Nonresponse*. New York: Wiley & Sons.

Bethlehem, J. (2010). Selection Bias in Web Surveys. *International Statistical Review, 78*(2), 161-188. doi:10.1111/j.1751-5823.2010.00112.x

Biemer, P. P. (2010). Total Survey Error: Design, Implementation, and Evaluation. *Public Opinion Quarterly, 74*(5), 817-848. doi:10.1093/poq/nfq058

Biemer, P. P., & Lyberg, L. E. (2003). *Introduction to Survey Quality*: John Wiley & Sons.

Billiet, J. (2003). Cross-Cultural Equivalence with Structural Equation Modeling. In J. Harkness, F. J. R. Van de Vijer, & P. Mohler (Eds.), *Cross-Cultural Survey Methods* (pp. 247-264). Hoboken: Wiley & Sons.

Bjanesoy, L. L. (2019). Effects of the Refugee Crisis on Perceptions of Asylum Seekers in Recipient Populations. *Journal of Refugee Studies, 32*(1), 219-237. doi:10.1093/jrs/fey070

Bloch, A., & Levy, C. (1999). *Refugees, Citizenship and Social Policy in Europe*. Houndmills: Macmillan Press.

Blom, A. G., Jäckle, A., & Lynn, P. (2010). The Use of Contact Data in Understanding Differences in Unit Nonresponse. In J. Harkness, M. Braun, B. Edwards, T. P. Johnson, L. Lyberg, P. Mohler, B.-E. Pennell, & T. W. Smith (Eds.), *Survey Methods in Multinational, Multiregional, and Multicultural Contexts* (pp. 335-354). Hoboken: Wiley & Sons.

BMI. (2016). 890.000 Asylsuchende im Jahr 2015. Retrieved from https://www.bmi.bund.de/SharedDocs/pressemitteilungen/DE/2016/09/asylsuchende-2015.html, last visited: 02.04.2020

Boda, Z. (2018). Social Influence and Observed Race. *Sociological Science*(5), 29-57. doi:10.15195/v5.a3

Böhmer, A., Damm, A.-C., Degler, F., Dzaferovic, M., Goebel, S., Hofmann, C., . . . Reinhart, E. (2018). *Wege der Integration*. Tutzing: Akademie für politische Bildung.

Böke, K. (1996). Flüchtlinge und Vertriebene zwischen dem Recht auf die alte Heimat und der Eingliederung in die neue Heimat. In K. Böke, F. Liedtke, & M. Wengeler (Eds.), *Sprache, Politik, Öffentlichkeit. Politische Leitvokabeln in der Adenauer-Ära* (pp. 131-210). Berlin: de Gruyter.

Bourhis, R. Y., Moise, C.-L., Senécal, S., & Perreault, S. (1997). Immigration und Multikulturalismus in Kanada: Die Entwicklung eines interaktiven Akkulturationsmodells. In A. Mummendey & B. Simon (Eds.), *Identität und Verschiedenheit. Zur Sozialpsychologie der Identität* (pp. 63-107). Bern: Huber.

Bourhis, R. Y., Moise, L. C., Perreault, S., & Senécal, S. (1997). Towards an Interactive Acculturation Model: A Social Psychological Approach. *International Journal of Psychology, 32*(6), 369-386. doi:10.1080/002075997400629

Brücker, H., Rother, N., & Schupp, J. (2017). IAB-BAMF-SOEP Befragung von Geflüchteten 2016: Studiendesign, Feldergebnisse sowie Analysen zu schulischer und beruflicher Qualifikation, Sprachkenntnissen sowie kognitiven Potentialen. *DIW Berlin Politikberatung kompakt,* 123

Brücker, H., Croisier, J., Kosyakova, Y., Kröger, H., Pietrantuono, G., Rother, N., & Schupp, J. (2019). Language Skills and Employment Rate of Refugees in Germany Improving with Time. *DIW Weekly Report*(4/5/6), 49-61.

Brücker, H., Kunert, A., Mangold, U., Kalusche, B., Siegert, M., & Schupp, J. (2016). Geflüchtete Menschen in Deutschland - eine qualitative Befragung. *IAB-Forschungsbericht*(9).

Chamberlain, G. (1984). Panel Data. In Z. Griliches & M. D. Intriligator (Eds.), *Handbook of Econometrics* (Vol. 2, pp. 1247-1318). Amsterdamt: Elsevier.

Cho, Y. (2014). To Know Democracy Is to Love It: A Cross-National Analysis of Democratic Understanding and Political Support for Democracy. *Political Research Quarterly, 67*(3), 478-488. doi:10.1177/1065912914532721

Coleman, J. S. (2001). *Grundlagen der Sozialtheorie*. Cambridge: Harvard University Press.

Comanaru, R., & d'Ardenne, J. (2018). The Development of a Research Programme to Translate and Test the Personal Well-being Questions in Sylheti and Urdu. In D. Behr (Ed.), *Surveying the Migrant Population. Consideration of Linguistic and Cultural Issues*. Köln: GESIS.

Couper, M. P., & de Leeuw, E. (2003). Nonresponse in Cross-Cultural and Cross-National Surveys. In J. Harkness, F. J. R. Van de Vijer, & P. Mohler (Eds.), *Cross-Cultural Survey Methods* (pp. 157-178). Hoboken: Wiley & Sons.

Couper, M. P., & Peterson, G. J. (2016). Why Do Web Surveys Take Longer on Smartphones? *Social Science Computer Review, 35*(3), 357-377. doi:10.1177/0894439316629932

Crul, M., & Schneider, J. (2010). Comparative Integration Context Theory: Participation and Belonging in New Diverse European Cities. *Ethnic and Racial Studies, 33*(7), 1249-1268. doi:10.1080/01419871003624068

da Lomba, S. (2010). Legal Status and Refugee Integration: a UK Perspective. *Journal of Refugee Studies, 23*(4), 415-436. doi:10.1093/jrs/feq039

de Bruijne, M., & Wijnant, A. (2013). Comparing Survey Results Obtained via Mobile Devices and Computers: An Experiment With a Mobile Web Survey on a Heterogeneous Group of Mobile Devices Versus a Computer-Assisted Web Survey *Social Science Computer Review, 31*(4), 482-504. doi:10.1177/0894439313483976

de Leeuw, E. (1993). Mode Effects in Survey Research: A Comparison of Mail, Telephone, and Face to Face Surveys. *Bulletin of Sociological Methodology*(41), 3-19.

Delnevo, C. D., Gundersen, D. A., & Hagman, B. T. (2008). Declining Estimated Prevalence of Alcohol Drinking and Smoking among Young Adults Nationally: Artifacts of Sample Undercoverage? *American Journal of Epidemiology, 167*(1), 15-19. doi:10.1093/aje/kwm313

Deming, W. E. (1944). On Errors in Survey. *American Sociological Review, 9*(4), 359-369.

DStatis. (2019). Migration. Retrieved from https://www.destatis.de/EN/Themes/Society-Environment/Population/Migration/_node.html;jsessionid=BFADCFC031A1A49E2BA5ACB9DFDA3495.internet712, last visited: 02.04.2020

Dustmann, C., & Kirchkamp, O. (2002). The Optimal Migration Duration and Activity Choice after Remigration. *Journal of Development Economics, 67*(2), 315-372. doi:10.1016/S0304-3878(01)00193-6

Eckman, S., & Kreuter, F. (2018). Misreporting to Looping Questions in Surveys: Recall, Motivation and Burden *Survey Research Methods, 12*(1), 59-74. doi:10.18148/srm/2018.v12i1.7168

Eisenstadt, S. N. (1954). *The Absporption of Immigrants. A Comparative Study Based Mainly on the Jewish Community in Palestine and the State of Israel*. London: Routledge.
Esser, H. (1980). *Aspekte der Wanderungssoziologie. Assimilation und Integration von Wanderern, ethnischen Gruppen und Minderheiten. Eine handlungstheoretische Analyse.* Darmstadt: Luchterhand.
Esser, H. (1993). *Soziologie. Allgemeine Grundlage*. Frankfurt am Main: Campus.
Esser, H. (1999). Inklusion, Integration und ethnische Schichtung. *Journal für Konflikt und Gewaltforschung, 1*(1), 5-34.
Esser, H. (2006). *Sprache und Integration. Die sozialen Bedingungen und Folgen des Spracherwerbs von Migranten*. Frankfurt am Main: Campus
Esser, H. (2009). Pluralisierung oder Assimilation ? Effekte der multiplen Inklusion auf die Integration von Migranten. *Zeitschrift für Soziologie, 38*(5), 358-378.
Eurobarometer. (2020). Eurobarometer. The EP and the Expectations of European Citizens. Retrieved from https://www.europarl.europa.eu/at-your-service/en/be-heard/eurobarometer, last visited: 02.04.2020
Eurostat. (2020). Terrritorial typologies manual. Retrieved from https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Territorial_typologies_manual, last visited: 02.04.2020
Farwick, A. (2011). The Effect of Ethnic Segregation on the Process of Assimilation. In M. Wingens, M. Windzio, H. de Valk, & C. Aybek (Eds.), *A Life-Course Perspective on Migration and Integration* (pp. 239-259). Dordrecht: Springer Science+Business.
Ferguson, T. S. (1996). *A Course in Large Sample Theory*. New York: Routledge.
Finkel, S. E., & Smith, A. E. (2011). Civic Education, Political Discussion, and the Social Transmission of Democratic Knowledge and Values in a New Democracy: Kenya 2002. *American Journal of Political Science, 55*(2), 417-435.
Foroutan, N. (2016). Postmigrantische Gesellschaften. In H. U. Brinkmann & M. Sauer (Eds.), *Einwanderungsgesellschaft Deutschland: Entwicklung und Stand der Integration* (pp. 227-254). Wiesbaden: Springer VS.
Foroutan, N., Kalter, F., Canan, C., & Zajak, S. (2019). *Ost-Migrantische Analogien I. Konkurrenz um Anerkennung*. DeZIM
Frick, J. R., & Grabka, M. M. (2010). Item Nonresponse and Imputation of Annual Labor Income in Panel Surveys from a Cross-National Perspective. In J. Harkness, M. Braun, B. Edwards, T. P. Johnson, L. Lyberg, P. Mohler, B.-E. Pennell, & T. W. Smith (Eds.), *Survey Methods in Multinational, Multiregional, and Multicultural Contexts* (pp. 355-374). Hoboken: Wiley & Sons.
Gambaro, L., Kreyenfeld, M., Schacht, D., & Spieß, C. K. (2018). Refugees in Germany with Children Still Living Abroad Have Lowest Life Satisfaction. *DIW Weekly Report*(42), 415-425.
Gerhards, J., & Hans, S. (2009). From Hasan to Herbert: Name-Giving Patterns of Immigrant Parents between Acculturation and Ethnic Maintenance. *American Journal of Sociology, 114*(4), 1102-1128. doi:10.1086/595944
Gerhards, J., & Kämpfer, S. (2017). Symbolische Grenzen und die Grenzarbeit von Migrantinnen und Migranten. Ein Typologisierungsvorschlag am Beispiel des Umgangs mit Vornamen. *Zeitschrift für Soziologie, 46*(5), 303-325. doi:10.1515/zfsoz-2017-1017
Giesselmann, M., & Windzio, M. (2012). *Regressionsmodelle zur Analyse von Paneldaten*. Wiesbaden: Springer VS.

Göbel, J., Grabka, M. M., Liebig, S., Kroh, M., Richter, D., Schröder, C., & Schupp, J. (2019). The German Socio-Economic Panel (SOEP). *Jahrbücher für Nationalökonomie und Statistik, 239*(2), 345-360.

Goebel, S. (2019). Push back and pull down: Arbeit als Steuerungsinstrument der Asylpolitik. In A. David, M. Evans, & I. Hamburg (Eds.), *Migration und Arbeit (Revisited) - Herausforderungen, Problemlagen, Gestaltungsinstrumente* (pp. 105-128). Opladen: Barbara Budruch.

Goerman, P., Meyers, M., & Trejo, Y. G. (2018). The Place of Expert Review in Translation and Questionnaire Evaluation for Hard-to-Count Populations in National Surveys. In D. Behr (Ed.), *Surveying the Migrant Population. Consideration of Linguistic and Cultural Issues* (pp. 29-42). Köln: GESIS.

Gordon, M. M. (1964). *Assimilation in American Life. The Role of Race, Religion, and National Origins.* New York: Oxford University Press.

Granato, N., & Kalter, F. (2001). Die Persistenz ethnischer Ungleichheit auf dem deutschen Arbeitsmark. Diskriminierung oder Unterinvestition in Humankapital? *Kölner Zetischrift für Soziologie und Sozialpsychologie, 53*(3), 497-520. doi:10.1007/s11577-001-0076-4

Green, M., King, E., & Fischer, F. (2019). Acculturation, Social Support and Mental Health Outcomes among Syrian Refugees in Germany *Journal of Refugee Studies, online first*. doi:10.1093/jrs/fez095

Greve, J., Schnabel, A., & Schützeichel, R. (2008). Das Makro-Mikro-Makro-Modell der soziologischen Erklärung - zur Einleitung. In J. Greve, A. Schnabel, & R. Schützeichel (Eds.), *Das Mikro-Makro-Modell der soziologischen Erklärung* (pp. 7-17). Wiesbaden: VS Verlag für Sozialwissenschaften.

Groves, R. M. (1989). *Survey Erros and Survey Costs*. New York: Wiley & Sonst.

Groves, R. M. (2006). Nonresponse Rates and Nonresponse Bias in Household Surveys. *Public Opinion Quarterly, 70*(5), 646-675. doi:10.1093/poq/nfl033

Groves, R. M., & Couper, M. P. (2012). *Nonresponse in Household Interview Surveys*. New York: John Wiley & Sons.

Groves, R. M., Fowler, F., Couper, M. P., Singer, E., & Tourangeau, R. (2004). *Survey Methodology*. New York: Wiley.

Groves, R. M., & Lyberg, L. (2010). Total Survey Error: Past, Present, and Future. *Public Opinion Quarterly, 74*(5), 849-879. doi:10.1093/poq/nfq065

Häder, S., & Gabler, S. (2003). Sampling and Estimation. In J. A. Harkness, F. J. R. Van de Vijer, & P. P. Mohler (Eds.), *Cross-Cultural Survey Methods* (pp. 117-136). Hoboken: John Wiley & Sons.

Hainmueller, J., & Hangartner, D. (2013). Who Gets a Swiss Passport? A Natural Experiment in Immigrant Discrimination. *American Political Science Review, 107*(1), 159-187. doi:10.1017/S0003055412000494

Hainmueller, J., Hangartner, D., & Pietrantuono, G. (2015). Naturalization Fosters the Long-Term Political Integration of Immigrants. *Proceedings of the National Academy of Sciences, 112*(41), 12651-12656. doi:10.1073/pnas.1418794112

Hainmueller, J., Hangartner, D., & Pietrantuono, G. (2017). Catalyst or Crown: Does Naturalization Promote the Long-Term Social Integration of Immigrants. *American Political Science Review, 111*(2), 256-276. doi:10.1017/S0003055416000745

Hanefeld, U. (1987). *Das Sozio-ökonomische Panel*. Frankfurt am Main: Campus Verlag.

Hans, S. (2010). *Assimilation oder Segregation? Anpassungsprozesse von Einwanderern in Deutschland*. Wiesbaden: VS Verlag für Sozialwissenschaften.

Hans, S. (2016). Theorien der Integration von Migranten– Stand und Entwicklung. In H. U. Brinkmann & M. Sauer (Eds.), *Einwanderungsgesellschaft Deutschland* (pp. 23-50). Wiesbaden: Springer VS.

Hansen, M. H., Hurwitz, W. N., & Madow, W. G. (1953). *Sample Survey Methods and Theory*. New York: John Wiley & Sons.

Hansen, S. E., Benson, G., Bowers, A., Pennell, B.-E., Lin, Y.-c., Duffey, B., . . . Hibben, K. C. (2016). *Survey Quality. Guidelines for Best Practice in Cross-Cultural Surveys*. Ann Arbor: Survey Research Center.

Harkness, J. A. (2003). Questionnaire Translation. In J. A. Harkness, F. J. R. Van de Vijer, & P. P. Mohler (Eds.), *Cross-Cultural Survey Methods* (pp. 35-56). Hoboken: John Wiley & Sons.

Harkness, J. A., Van de Vijer, F. J. R., & Mohler, P. P. (2003). *Cross Cultural Survey Methods*. Hoboken: Johne Wiley & Sons.

Harter, R., Eckman, S., English, N., & O'Muircheartaigh, C. (2010). Applied Sampling for Large-Scale Multistage Area Probability Designs. In P. V. Marsden & J. D. Wright (Eds.), *Handbook of Survey Research* (pp. 169-195). Bingley: Emerald Group Publishing.

Heckman, J. J. (1979). Sample Selection Bias as a Specification Error. *Econometrica, 47*(1), 153-161. doi:10.2307/1912352

Helle, H. J. (1985). *Macro-sociological Theory*. London: SAGE Publications.

Huntington, S. P. (1993). The Clash of Civilisations? *Foreign Affairs, 72*(3), 22-49. doi:10.2307/20045621

IAB. (2019). *IAB-BAMF-SOEP Survey of Refugees*.

Ingwersen, K., & Thomsen, S. L. (2019). The Immigrant-Native Wage Gap in Germany Revisited. *IZA Discussion Paper Series*(12358).

Jacobsen, J. (2018). Language Barriers During the Fieldwork of the IAB-BAMF-SOEP Survey of Refugees in Germany. In D. Behr (Ed.), *Surveying the Migrant Population: Consideration of Linguistic and Cultural Issues*. Köln: GESIS.

Jacobsen, J. (2019). An Investment in the Future: Institutional Aspects of Credential Recognition of Refugees in Germany. *Journal of Refugee Studies, online first*. doi:10.1093/jrs/fez094

Jacobsen, J., Jaschke, P., Krieger, M., Kroh, M., Legewie, N., & Löbel, L.-M. (forthcoming). Bringing Together Community Organizations and Social Science Research – The "Mentoring of Refugees" Project. *Canadian Diversity*.

Jacobsen, J., Krieger, M., & Legewie, N. (forthcoming). Labor Market Participation of Refugees in Germany: Legal Context and Individual-Level Factors. In A. Korntheuer, D. B. Maehler, P. Pritchard, & L. Wilkinson (Eds.), *Refugees in Canada and Germany: Responses in Policy and Practice* (pp. 173-185). Köln: GESIS.

Jacobsen, J., Kroh, M., Kühne, S., Scheible, J. A., Siegers, R., & Siegert, M. (2019). Supplementary of the IAB-BAMF-SOEP Survey of Refugees in Germany (M5) 2017. *SOEP Survey Papers*(605).

Jacobsen, J., Kroh, M., Legewie, N., & Salikutluk, Z. (2018). Arbeitsmarktpotenzial, Qualifikationsniveau, berufliche Perspektiven Geflüchteter. In H. Floeting, G. Kirchhoff, H. Scheller, & J. Schneider (Eds.), *Zuwanderung und Integration von Geflüchteten in Kommunen: Der lange Weg vom Ankommen zum Bleiben* (pp. 99-105). Berlin: Deutsches Institut für Urbanistik.

Jagers, P. (1986). Post-Stratification Against Bias in Sampling. *International Statistical Review, 54*(2), 159-167. doi:10.2307/1403141

Jann, B., Jerke, J., & Krumpal, I. (2011). Asking Sensitive Questions Using the Crosswise Model: An Experimental Survey Measuring Plagiarism. *Public Opinion Quarterly, 76*(1), 32-49. doi:10.1093/poq/nfr036
Jenkins, R. (2014). Time to Move Beyond Boundary Making? *Ethnic and Racial Studies, 37*(5), 809-813. doi:10.1080/01419870.2013.871308
Johnson, T. P., & Van de Vijer, F. J. R. (2003). *Social Desirability in Cross-Cultural Research*. Hoboken: Wiley & Sons.
Juran, J. M., & Gryna, F. M. (1993). *Quality Planning and Analysis*. New York: McGraw-Hill.
Kalgraff Skjak, K., & Harkness, J. A. (2003). Data Collection Methods. In J. A. Harkness, F. J. R. Van de Vijer, & P. P. Mohler (Eds.), *Cross-Cultural Survey Methods* (pp. 179-194).
Kalter, F. (2006). Auf der Suche nach einer Erklärung für die spezifischen Arbeitsmarktnachteile von Jugendlichen türkischer Herkunft. *Zeitschrift für Soziologie, 35*(2), 144-160. doi:10.1515/zfsoz-2006-0204
Kalter, F. (2008). *Migration und Integration*. Wiesbaden: VS Verlag für Sozialwissenschaften.
Kalter, F., & Kogan, I. (2014). Migrant Networks and Labor Market Integration of Immigrants from the Former Soviet Union in Germany *Social Forces, 92*(4), 1435-1456. doi:10.1093/sf/sot155
Kish, L. (1965). *Survey Sampling*. New York: John Wiley & Sons.
Kogan, I. (2016). Integration Policies and Immigrants' Labor Market Outcomes in Europe. *Sociological Science*(3), 335-358. doi:10.15195/v3.a16
Kogan, I., Kalter, F., Liebau, E., & Coher, Y. (2011). Individual Resources and Structural Constraints in Immigrants' Labour Market Integration. In M. Wingens, M. Windzio, H. de Valk, & C. Aybek (Eds.), *A Life-Course Perspective on Migration and Integration* (pp. 75-100). Dordrecht: Springer.
Kreft, I., & de Leeuw, J. (1998). *Introducing Multilevel Modeling*. London: Sage Publications.
Kreisberg, A. N. (2019). Starting Points: Divergent Trajectories of Labor Market Integration among U.S. Lawful Permanent Residents *Social Forces, 98*(2), 849-884. doi:10.1093/sf/soy128
Krieger, M. (2019). Tied and Troubled: Revisiting Tied Migration and Subsequent Employment. *Journal of Marriage and Family, online first*. doi:10.1111/jomf.12620
Kroh, M., Käppner, K., & Kühne, S. (2014). Sampling, Nonresponse, and Weighting in the 2011 and 2012 Refreshment Samples J and K of the Socio-Economic Panel. *SOEP Survey Papers*(260).
Kroh, M., Kühne, S., Goebel, J., & Preu, F. (2015). The 2013 IAB-SOEP Migration Sample (M1): Sampling Design and Weighting Adjustment. *SOEP Survey Papers*(271).
Kroh, M., Kühne, S., Jacobsen, J., Siegert, M., & Siegers, R. (2017). Sampling, Nonresponse, and Integrated Weighting of the 2016 IAB-BAMF-SOEP Survey of Refugees (M3/M4) - Revised Version. *SOEP Survey Papers*(477).
Krosnick, J. A. (1991). Response Strategies for Coping with the Cognitive Demands of Attitude Measures in Surveys. *Applied Cognitive Psychology, 5*(3), 213-236. doi:10.1002/acp.2350050305
Krumpal, I. (2013). Determinants of Social Desirability Bias in Sensitive Surveys: a Literature Review. *Quality & Quantity, 47*(4), 2025-2047. doi:10.1007/s11135-011-9640-9

Kruse, H., & Kroneberg, C. (2019). More Than a Sorting Machine: Ethnic Boundary Making in a Stratified School System. *American Journal of Sociology, 125*(2), 431-484. doi:10.1086/705561
Kubiak, D., & Foroutan, N. (2018). Ausschluss und Abwertung: Was Muslime und Ostdeutsche verbindet. *Blätter für deutsche und internationale Politik*(7), 93-102.
Kühne, S. (2018). From Strangers to Acquaintances? Interviewer Continuity and Socially Desirable Responses in Panel Surveys *Survey Research Methods, 12*(2), 121-146. doi:10.18148/srm/2018.v12i2.7299
Kühne, S., Jacobsen, J., & Kroh, M. (2019). Sampling in Times of High Immigration: The Survey Process of the IAB-BAMF-SOEP Survey of Refugees. *Survey Methods: Insights from the Field*. doi:10.13094/SMIF-2019-00005
Legewie, N., Jaschke, P., Krieger, M., Kroh, M., Löbel, L.-M., & Schacht, D. (2019). Integrating Randomized Controlled Field Trials into (Existing) Panel Surveys: The "Mentoring of Refugees" Study. *Briefs on Methodological, Ethical and Epistemological Issues*(7).
Leibold, J. (2006). *Immigranten zwischen Einbürgerung und Abwanderung. Eine empirische Studie zur bindenden Wirkung von Sozialintegration.* (Promotion). Georg-August-Universität, Göttingen.
Leibold, J., & Kühnel, S. (2016). Migranten und Einheimische - Welche wechselseitigen Wahrnehmuingen haben sich im Verlauf der Zeit durchgesetzt? In H. U. Brinkmann & M. Sauer (Eds.), *Einwaunderungsgesellschaft Deutschland. Entwicklung und Stand der Integration* (pp. 311-338). Wiesbaden: Springer VS.
Liebau, E., Humpert, A., & Schneiderheinze, K. (2018). Wie gut funktioniert das Onomastik-Verfahren? Ein Test am Beispiel des SOEP-Datensatzes. *SOEPpapers*(976).
Liebau, E., & Schupp, J. (2011). Considering Emigration: German University Graduates are Moving Abroad - but Only Temporarily. *DIW Weekly Report, 7*.
Liebau, E., & Tucci, I. (2015). Migrations- und Integrationsforschung mit dem SOEP von 1984 bis 2012: Erhebung, Indikatoren und Potenziale. *SOEP Survey Papers*(270).
Löbel, L.-M. (2019). Family Separation and Refugee Mental Health: a Network Perspective. *Social Networks, online first*. doi:10.1016/j.socnet.2019.08.004
Loosveldt, G., Pickery, J., & Billiet, J. (2002). Item Nonresponse as a Predictor of Unit Nonresponse in a Panel Survey. *Journal of Official Statistics, 18*(4), 545-557.
Lugtig, P., Lensvel-Mulders, G., Frerichs, R., & Greven, A. (2011). Estimating Nonresponse Bias and Mode Effects in a Mixed-Mode Survey *International Journal of Market Research, 53*(5), 669-686. doi:10.2501/IJMR-53-5-669-686
Luhmann, N. (1995). Inklusion und Exklusion. In N. Lumann (Ed.), *Die Soziologie und der Mensch* (pp. 237-264). Opladen: Westdeutscher Verlag.
Lyberg, L. (2012). Survey Quality. *Survey Methodology, 38*(2), 107-130.
Mägi, M., van Ham, M., Leetmaa, K., & Tammaru, T. (2018). The Neighbourhood Context and Changes in Self-Reported Ethnic Identity. *Journal of Ethnic and Migration Studies, online first*. doi:10.1080/1369183X.2018.1547634
Martin, P., & Lynn, P. (2011). The Effects of Mixed Mode Survey Designs on Simple and Complex Analyses. *ISER Working Paper Series, 2018*(28).
Mau, S. (2019). *Lütten Klein. Leben in der ostdeutschen Transformationsgesellschaft*. Berlin: Suhrkamp.
McFarland, S. G. (1981). Effects of Question Order on Survey Responses. *Public Opinion Quarterly, 45*(2), 208-215. doi:10.1086/268651

Miyamoto, A., Seuring, J., & Kristen, C. (2018). Immigrant Students' Achievements in Light of Their Educational Aspirations and Academic Motivation. *Journal of Ethnic and Migration Studies, online first*. doi:10.1080/1369183X.2018.1538772

Mohler, P., Dorer, B., de Jong, J., & Hu, M. (2016). *Translation. Guidelines for Best Practice in Cross-Cultural Surveys.* . Ann Arbor: Survey Research Center.

Morganstein, D., & Marker, D. A. (1997). Continuous Quality Improvement in Statistical Agencies. In L. Lyberg, P. Biemer, M. Collins, E. De Leeuw, C. Dippo, N. Schwarz, & D. Trewin (Eds.), *Survey Measurement and Process Quality*. Hoboken: Wiley & Sons.

Münkler, H., & Münkler, M. (2016). *Die neuen Deutschen. Ein Land vor seiner Zukunft*. Berlin: Rowohlt.

Mustillo, S. A., Lizardo, O. A., & McVeigh, R. M. (2018). Editors' Comment: A Few Guidelines for Quantitative Submissions. *American Sociological Review, 83*(6), 1281-1283. doi:10.1177/0003122418806282

Nauck, B. (2008). Akkulturation: Theoretiscge Ansätze und Perspektiven der Psychologie und Soziologie. In F. Kalter (Ed.), *Migration und Integration* (pp. 108-132). Wiesbaden: VS Verlag für Sozialwissenschaften.

Oltmer, J. (2016). Europäische und deutsche Migrationsverhältnisse im 19. und 20. Jahrhundert. In H. U. Brinkmann & M. Sauer (Eds.), *Einwanderungsgesellschaft Deutschland. Stand und Entwicklung der Integration* (pp. 51-98). Wiesbaden: Springer VS.

Ong, A. D., & Weiss, D. J. (2000). The Impact of Anonymity on Responses to Sensitive Questions. *Journal of Applied Social Psychology, 30*(8), 1691-1708. doi:10.1111/j.1559-1816.2000.tb02462.x

Park, R. E. (1950). *Race and Culture*. Glencoe: Free Press.

Park, R. E., & Burgess, E. W. (1921). *Introduction to the Science of Sociology*. Chicago: University of Chicago Press.

Pennell, B.-E., Harkness, J. A., Levenstein, R., & Quaglia, M. (2010). Challenges in Cross-National Data Collection. In J. Harkness, M. Braun, B. Edwards, T. P. Johnson, L. Lyberg, P. Mohler, B.-E. Pennell, & T. W. Smith (Eds.), *Survey Methods in Multinational, Multiregional, and Multicultural Contexts* (pp. 269-298). Hoboken: Wiley & Sons.

Pennell, B.-E., Hibben, K. C., Lyberg, L. E., Mohler, P., & Worku, G. (2017). A Total Survey Error Perspective on Surveys in Multinational, Multiregional, and Multicultural Contexts. In P. P. Biemer, E. de Leeuw, S. Eckman, B. Edwards, F. Kreuter, L. E. Lyberg, N. C. Tucker, & B. T. West (Eds.), *Total Survey Error in Practice*. Hoboken: John Wiley & Sons.

Perlmann, J., & Waldinger, R. (1997). Second Generation Decline? Children of Immigrants, past and Present - A Reconsideration. *International Migration Review, 31*(4), 893-922. doi:10.1177/019791839703100405

Phinney, J. S., & Ong, A. D. (2007). Conceptualization and Measurement of Ethnic Identity: Current Status and Future Directions. *Journal of Conseling Psychology, 54*(3), 271-281. doi:10.1037/0022-0167.54.3.271

Piekut, A. (2019). Survey Nonresponse in Attitudes Towards Immigration in Europe. *Journal of Ethnic and Migration Studies, online first*. doi:10.1080/1369183X.2019.1661773

Portes, A., & Rumbaut, R. G. (2001). *Legacies: The Story of Immigrant Second Generation*. Berkeley: University of California Press.

Portes, A., & Zhou, M. (1993). The New Second Generation: Segmented Assimilation and Its Variants. *Annals of the American Academy of Political and Social Science, 530*, 74-96.
Pries, L. (2016). *Migration und Ankommen*. Frankfurt am Main: Campus.
Pritchard, P., Maehler, B. B., Pötzschke, S., & Ramos, H. (2019). Integrating Refugee Children and Youth: A Scoping Review of English and German Literature. *Journal of Refugee Studies, 32*(1), 194-208. doi:10.1093/jrs/fez024
Ramos, M. R., Cassidy, C., Reicher, S., & Haslam, S. A. (2016). A Longitudinal Study of the Effects of Discrimination on the Acculturation Strategies of International Students. *Journal of Cross-Cultural Psychology, 47*(3), 401-420. doi:10.1177/0022022116628672
Reckwitz, A. (2019). *The Society of Singularities [translated version of the 2017 published book in German]*. Cambridge: Polity Press.
Rendtel, U. (1995). *Lebenslagen im Wandel: Panelausfälle und Panelrepräsentativität*. Frankfurt am Main: Campus Verlag.
Sakshaug, J., Tutz, V., & Kreuter, F. (2013). Placement, Wording, and Interviewers: Identifying Correlates of Consent to Link Survey and Administrative Data *Survey Research Methods, 7*(2), 133-144. doi:10.18148/srm/2013.v7i2.5395
Salikutluk, Z. (2016). Why Do Immigrant Students Aim High?Explaining the Aspiration–Achievement Paradoxof Immigrants in Germany. *European Sociological Review, 32*(5), 581-592. doi:10.1093/esr/jcw004
Sandor, S. (2017). Teilhabe, Behinderung, Zuwanderung. *Zeitschrift für Inklusion, 25*(4), 226-231.
Schupp, J., & Wagner, G. G. (1995). Die Zuwanderer-Stichprobe des Sozio-Ökonomischen Panels (SOEP). *Vierteljahreshefte zur Wirtschaftsforschung, 64*(1), 16-25.
Schwarz, N. (2003). *Culture-Sensitive Context Effects: A Challenge for Cross-Cultural Surveys*. Hoboken: Wiley & Sons.
Schwarz, N., Park, D. C., Knäuper, B., & Sudman, S. (1999). *Cognition, Aging, and Self-reports*. Washington: Psychology Press.
Seifert, W. (2012). Geschichte der Zuwanderung nach Deutschland nach 1950. Retrieved from http://www.bpb.de/politik/grundfragen/deutsche-verhaeltnisse-eine-sozialkunde/138012/geschichte-der-zuwanderung-nach-deutschland-nach-1950?p=all, last visited: 02.04.2020
Shen, J., & Kogan, I. (2019). Immigrants' Relative Income and Life Satisfaction: Comparison Groups from a Multi-Generational Perspective. *Acta Sociologica, online first*. doi:10.1177/0001699319859397
Shils, E. (1975). *Center and Periphery: Essays in Macrosociology*. Chicago: University of Chicago Press.
Smith, T. W. (2011). Refining the Total Survey Error Perspective. *International Journal of Public Opinion Research, 23*(4), 464-484. doi:10.1093/ijpor/edq052
SOEP. (2019). *Sozio-oekonomisches Panel (SOEP), Daten der Jahre 1984-2017*.
Spörlein, C., & Kristen, C. (2018). Educational Selectivity and Language Acquisition among Recently Arrived Immigrants. *International Migration Review, 53*(4), 1148-1170. doi:10.1177/0197918318798343
Stark, T. H., Silber, H., Krosnick, J. A., Blom, A. G., Aoyagi, M., Belchior, A., . . . Yu, R.-r. (2018). Generalization of Classic Question Order Effects Across Cultures. *Sociological Methods & Research, online first*. doi:10.1177/0049124117747304
Strang, A., & Ager, A. (2010). Refugee Integration: Emerging Trends and Remaining Agendas. *Journal of Refugee Studies, 23*(4), 589-607. doi:10.1093/jrs/feq046

Tabib-Calif, Y., & Lomsky-Feder, E. (2014). Symbolic Boundary Work in Schools: Demarcating and Denying Ethnic Boundaries. *Anthropology & Education, 45*(1), 22-38. doi:10.1111/aeq.12045

Tajfel, H. (1981). *Human Groups and Social Categories*. Cambridge: Cambridge University Press.

Tajfel, H., & Turner, J. C. (2004). The Social Identity Theory of Intergroup Behavior. In J. T. Jost & J. Sidanius (Eds.), *Key readings in social psychology. Political psychology: Key readings* (pp. 276-293). New York: Psychology Press.

Tayi, G. K., & Ballou, D. P. (1998). Examining Data Quality. *Communications of the ACM, 41*(2), 54-57.

Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). *The Psychology of Survey Response*. Cambridge: Cambridge University Press.

Tourangeau, R., & Yan, T. (2007). Sensitive Questions in Surveys. *Psychological Bulletin, 133*(5), 859-883. doi:10.1037/0033-2909.133.5.859

Tucci, I. (2011). National Context and Logic of Social Distancing: Children of Immigrants in France and Germany. In M. Wingens, M. Windzio, H. de Valk, & C. Aybek (Eds.), *A Life-Course Perspective on Migration and Integration* (pp. 142-164). Dordrecht: Springer Science+Business.

UNHCR. (1951). Convention and Protocol Relating to the Status of Refugees. Retrieved from https://www.unhcr.org/3b66c2aa10, last visited: 02.04.2020

Valtonen, K. (2004). From the Margin to the Mainstream: Conceptualizing Refugee Settlement Processes. *Journal of Refugee Studies, 17*(1), 70-96. doi:10.1093/jrs/17.1.70

Van de Vijer, F. J. R. (2003). Bias and Equivalence: Cross-Cultural Perspectives. In J. Harkness, F. J. R. Van de Vijer, & P. Mohler (Eds.), *Cross-Cultural Survey Methods* (pp. 143-156). Hoboken: Wilex & Sons.

Wagner, G., Schupp, J., & Rendtel, U. (1994). Das Sozio-oekonomische Panel (SOEP) - Methoden der Datenproduktion und -aufbereitung im Längsschnitt. In R. Hauser, N. Ott, & G. Wagner (Eds.), *Mikroanalytische Grundlagen der Gesellschaftspolitik* (pp. 70-112). Berlin: Akademie Verlag.

Wals, S. C., & Rudolph, T. J. (2018). Lost in Acculturation? Premigratory Exposure to Democracy and Immigrants' Political Trust in the United States. *International Journal of Public Opinion Research, 31*(1), 70-92. doi:10.1093/ijpor/edy002

Walther, L., Fuchs, L. M., Schupp, J., & Scheve, C. v. (2020). Living Conditions and the Mental Health and Wellbeing of Refugees: Evidence from a LargeScale German Survey. *Journal of Immigrant and Minortiy Health, online first*. doi:10.1007/s10903-019-00968-5

Walther, L., Kröger, H., Tibubos, A. N., Tam Ta, T. M., von Scheve, C., Schupp, J., . . . Bajbouj, M. (2019). Psychological Distress among Refugees in Germany – a Representative Study on Individual and Contextual Risk Factors and the Potential Consequences of Poor Mental Health for Integration in the Host Country. *SOEPpapers*(1053).

Waters, M. C., & Jiménez, T. R. (2005). Assessing Immigrant Assimilation: New Empirical and Theoretical Challenges. *Annual Review of Sociology, 31*, 105-125. doi:10.1146/annurev.soc.29.010202.100026

Weichbold, M. (2009). Zur Bestimmung und Sicherung der "Qualität" von Umfragen. In M. Weichbold, J. Bacher, & C. Wolf (Eds.), *Umfrageforschung. Herausforderungen und Grenzen* (pp. 553-570). Wiesbaden: VS Verlag für Sozialwissenschaften.

West, B. T., & Blom, A. G. (2017). Explaining Interviewer Effects: A Research Synthesis. *Journal of Survey Statistics and Methodology, 5*(2), 175-211. doi:10.1093/jssam/smw024

Wimmer, A. (2008). The Making and Unmaking of Ethnic Boundaries: A Multilevel Process Theory. *American Journal of Sociology, 114*(4), 970-1022. doi:10.1086/522803

Wimmer, A. (2009). Herder's Heritage and the Boundary-Making Approach. Studying Ethnicity in Immigrant Societies. *Sociological Theory, 27*(3), 244-270. doi:10.1111/j.1467-9558.2009.01347.x

Wimmer, A. (2013). *Ethnic Boundary Making. Institutions, Power, Networks*. Oxford: Oxford University Press.

Wolfsteller, R., & Rädel, J. (2019). Über die Fallstricke sozialwissenschaftlicher Identitäts- und Einstellungsforschung: Ziele, Potenziale und Kritik der DeZIM-Studie »Ost-Migrantische Analogien I«. *Leviathan, 47*(4), 428-444. doi:10.5771/0340-0425-2019-4-428

Wolter, K. M. (2007). *Introduction to Variance Estimation*. Wiesband: Springer-Verlag New York.

Wrohlich, K., & Zucco, A. (2017). Gender Pay Gap innerhalb von Berufen variiert erheblich. *DIW Wochenbericht*(43), 955-961.

Yan, T., & Curtin, R. (2010). The Relation Between Unit Nonresponse and Item Nonresponse: A Response Continuum Perspective. *International Journal of Public Opinion Research, 22*(4), 535-551. doi:10.1093/ijpor/edq037

Yan, T., & Tounrangeau, R. (2008). Fast Times and Easy Questions: The Effects of Age,Experience and Question Complexity on Web SurveyResponse Times. *Applied Cognitive Psychology, 22*(1), 51-68. doi:10.1002/acp.1331

Yoon, E., Chang, C.-T., Kim, S., & Clawson, A. (2013). A Meta-Analysis of Acculturation/Enculturation and Mental Health. *Journal of Counseling Psychology, 60*(1), 15-30. doi:10.1037/a0030652

Zhou, M. (1997). Segmented Assimilation: Issues, Controversies, and Recent Research on the New Second Generation. *International Migration Review, 31*(4), 975-1008.

Zimmermann, R., Kaimer, S., & Oberschachtsiek, D. (2007). Dokumentation des "Scientific Use Files der Integrierten Erwerbsbiographien" (IEBS-SUF V1) Version 1.0. *FDZ-Datenreport*(1).

4 Article 1: Sampling in Times of High Immigration

Preface

The purpose of quantitative research in the social sciences is to explain individual and collective behavior. Survey data that allow for inferences on a certain target population (e.g., general population, migrant cohorts, the unemployed) offer a source for such analyses. Usually, random samples are the most effective way to ensure generalizability because most mathematical and statistical tools of inferential statistics rely on the assumption that the data were sampled randomly. Therefore, much of the documentation of survey data explicitly focuses on demonstrating randomness in the sampling process. However, less attention is usually given to the choice of the sampling frame and its implications for generalizability. This is because the options for sampling frames are rather limited in general. But if the sampling frame does not cover the population of interest sufficiently, even random sampling and the most elaborate analytical methodology will not allow for inference.

With the increasing interest in data on refugees, the question of an adequate sampling frame for this population has attracted growing interest. At first glance, sampling recent refugees in Germany is straightforward because all foreign nationals who live in Germany for more than three months are listed in the Central Register of Foreigners (AZR). However, as the Federal Office for Migration and Refugees (BAMF), which administers the AZR, has acknowledged, during the peak period of refugee migration in 2015 and 2016, there was a time lag between crossing the border and being registered in the AZR.

The first paper in my dissertation takes this issue as an example and proposes a novel approach for dealing with time-lagged sampling frames: sequential sampling. Besides explaining the sampling procedure, the paper describes the weighting process and the implementation of translated field instruments.

Sampling in Times of High Immigration: The Survey Process of the IAB-BAMF-SOEP Survey of Refugees

The article is published and should be cited as follows:
Kühne, Simon; Jacobsen, Jannes; Kroh, Martin (2019) Sampling in Times of High Immigration: The Survey Process of the IAB-BAMF-SOEP Survey of Refugees. *Survey Methods: Insights from the Field.* DOI: 10.13094/SMIF-2019-00005[1]

Abstract

Over the course of 2013 to 2016, over one million asylum seekers arrived in Germany, around 890,000 of them in 2015 alone. The growing refugee population posed a major challenge for Germany's policy makers, civic administrators, and society at large, in finding new approaches to registration procedures, housing, and social and economic integration. To design policies and programs that meet these needs, government administrators, politicians, and the public require robust analyses of the accompanying social and demographic changes based on timely, valid, and reliable empirical data. Yet despite the urgent need for quantitative data on this target group, survey organizations and data collection agencies had little experience gaining access to the target population and approaching and surveying them effectively.

In late 2015, when the influx reached its peak, the Institute for Employment Research (IAB), the Migration, Integration and Asylum Research Center at the Federal Office for Migration and Refugees (BAMF-FZ), and the Socio-Economic Panel (SOEP) joined together in a cooperative longitudinal project to survey a nationwide random sample of refugee households in Germany: the IAB-BAMF-SOEP Survey of Refugees. In this paper, we summarize the sampling and fieldwork design as well as the challenges faced in the IAB-BAMF-SOEP Survey of Refugees. We discuss the sequential strategy applied for sampling recent refugees and asylum seekers who arrived in Germany, particularly in

[1] Please note that after publication, some households have been deleted from the scientific usefile due to inadequately conducted interviews (see Kroh et al, 2017). Results and conclusions, however, do not change.

2015, in such large numbers that proper registration was delayed, and in many cases their initial accommodations were only temporary. Moreover, the paper discusses alternative survey instruments introduced for the difficult-to-interview population of the IAB-BAMF-SOEP Survey of Refugees, including translated questionnaires and audio files.

1. Introduction

Over the course of 2013 to 2016, over one million asylum seekers arrived in Germany, around 890,000 of them in 2015 alone (see the press release BMI, 2016). The growing refugee population posed a major challenge for Germany's policy makers, civic administrators, and society at large, in finding new approaches to registration procedures, housing, and social and economic integration. To design policies and programs that meet these needs, government administrators, politicians, and the public require robust analyses of the accompanying social and demographic changes based on timely, valid, and reliable empirical data. Yet despite the urgent need for quantitative data on this target group, survey organizations and data collection agencies had little experience gaining access to the target population and approaching and surveying them effectively.

In late 2015, when the influx reached its peak, the Institute for Employment Research (IAB), the Integration and Asylum Research Center at the Federal Office for Migration and Refugees (BAMF-FZ), and the Socio-Economic Panel (SOEP) joined together in a cooperative longitudinal project to survey a nationwide random sample of refugee households in Germany: the IAB-BAMF-SOEP Survey of Refugees. Funding came primarily from the Federal Ministry of Labour and Social Affairs (BMAS) and the Federal Ministry of Education and Research (BMBF). In the first wave in 2016, a total of 3,336 households were interviewed, resulting in 4,527 face-to-face interviews with individual adult respondents. An enlargement sample in 2017 added an additional 1,519 households and 2,252 individuals. Together, the samples are representative of the population of refugees and asylum seekers who arrived in Germany between 2013 and 2016 and were registered in the Central Register of Foreigners by January 2017. The scientific use file of the data is made available by the SOEP Research Data Center to the scientific community (DOI: https://doi.org/10.5684/soep.v33.1) as well as the IAB Research Data Center (https://fdz.iab.de/en.aspx).

Refugees and asylum seekers living in Germany are entered into the Central Register of Foreigners ("Ausländerzentralregister", AZR, see von Gostomski & Pupeter, 2008), which is a national administrative list of individuals from foreign countries living in Germany. The register is maintained by a department of the Federal Office for Migration and Refugees (BAMF). Sampling the target population is relatively straightforward if one has

access to this register, which is essentially only available to BAMF staff. Although our cooperation with BAMF guaranteed access to the register, two factors complicated the sampling process.

First, particularly in 2015, the responsible authorities were unable to cope with the high numbers of incoming refugees, both in terms of registration at the border as well as in processing their applications for asylum and refugee status. This latter administrative procedure is necessary, however, for individuals to be entered into the Central Register of Foreigners and identified as refugees for sampling. According to official statistics provided by the German Ministry of the Interior (BMI, 2016) and the BAMF (BAMF, 2015, p. 2; 2018), not all individuals who migrated to Germany as potential refugees were actually identified by the AZR at times of the highest influx numbers in 2015. As the data by the BAMF indicates around 480k people were registered as asylum seekers while actually around 890k were estimated to be in the country. Hence, the register was plagued at times by considerable undercoverage of the target population of incoming refugees. However, during the year 2016 this undercoverage was mostly resolved (BAMF 2018). Moreover, the delay affected all asylum seekers, not only some subgroups, thus, unlikely to introduce systematic effects on sample composition. Nonetheless, in order to account for this, a sequential sampling strategy was implemented to add individuals at later points who were otherwise not covered by the frame.

A second factor complicating sampling was the shortage of public housing and the high residential mobility of refugees, particularly shortly after their arrival in Germany. Incoming refugees typically first go to a refugee reception facility, where they stay for a short time, then move into publicly provided refugee housing, and eventually into private housing. These steps often took place within a matter of weeks, and refugees sometimes have to move large distances from one municipality or federal state to the next, according to the "EASY" (*Erstverteilung der Asylbegehrenden*) quota system designed to facilitate the distribution of refugees across the federal states. As a consequence, it is sometimes difficult to keep track of sample members' current addresses.

In addition, surveying refugees in Germany entails challenges in the actual fieldwork and interviewing procedures. These relate to the design of fieldwork instruments, the training of interviewers, and nonresponse of sampled households.

In this paper, we summarize the sampling and fieldwork design as well as the challenges faced in the IAB-BAMF-SOEP Survey of Refugees. Section 2 describes our sampling strategy: we sampled specific sample tranches at different time points in a step-by-step process, combined with timely sampling of selected clusters, in which the time between sampling and initial contact was reduced to a week. Section 3 describes the procedure used to interview the sampled refugee population, which is just beginning to learn German after arriving in Germany: Besides being difficult to reach, they are in some cases difficult to interview as well. The paper at hand discusses the insights and practical experiences gained so far in conducting the IAB-BAMF-SOEP Survey of Refugees in Germany.

2. Sampling

The IAB-BAMF-SOEP Survey of Refugees consists of multiple subsamples. All subsamples were drawn from the Central Register of Foreigners. For each subsample, we used the same sampling procedure: a two-stage clustered disproportional stratified sampling design (see Kroh, Kühne, Jacobsen, Siegert, & Siegers, 2017 for details; and Lohr, 2010 for general survey sampling theory and applications). In a first step, we selected primary sampling units (PSUs) representing regional clusters of immigration offices. Here, we made use of the fact that each individual in the register is assigned to a local immigration office. These offices are located across Germany and maintain information on the individual administrative procedures and addresses of foreigners and refugees living in the area. PSUs were selected with replacement and in 16 strata representing federal states and differentiated by county type (rural vs. urban). In each cluster, secondary sampling units (SSUs)—the individuals—were selected based on a disproportional sampling scheme that ensured minimum sample sizes and thus allowed for meaningful comparisons between subgroups of refugees. We assigned varying sample probabilities depending upon an individual's country of origin, current legal status, age, as well as gender.

2.1 Sampling in Tranches

As mentioned above, the Central Register of Foreigners was unable to keep up with the influx of refugees and asylum seekers (also referred to as the "EASY" gap), and thus, facing problems due to undercoverage. Moreover, asylum seekers and refugees are a

highly mobile target population, especially shortly after arrival. We chose to sample a total of six different "tranches" to address these issues. These tranches were sampled at four consecutive points in time using updated versions of the register. Moreover, later tranches not only focused on more recent arrivals to Germany, but also on refugees who had arrived earlier but appeared in the register late.

The scientific use file of the IAB-BAMF-SOEP Survey of Refugees documents the six tranches as three subsamples with somewhat different target populations (M3, M4, and M5), which are designed to be used jointly. The names result from the SOEP's standard procedure, where different subsamples are named in alphabetic order. Samples M1 and M2 are samples of migrants to Germany that existed prior to the IAB-BAMF-SOEP Survey of Refugees.

As with all other existing SOEP subsamples, M3, M4, and M5 are based on a household concept, according to which every (adult) household member is interviewed. Individuals selected from the register thus represent what are known as "anchor respondents": these are the household members with whom the field agency makes initial contact. They then add the rest of the household by interviewing each household member 18 years or older and collecting proxy information on children and adolescents. Design and household nonresponse weighting procedures allow for representative analyses at both household and individual level.

Table 1 displays characteristics of the six sample tranches. Asylum seekers and refugees who arrived in Germany between January 2013 and January 2016 were the target population for subsamples M3 and M4. Subsample M5 is both a refresher of the M3/M4 population as well as an enlargement sample of asylum seekers and refugees who arrived in Germany between February 2016 and the end of December 2016. A total of four versions of the Central Register of Foreigners were used to address potential gaps in coverage of the population due to the lag in registration. For instance, Sample M3-2 included only those anchor respondents who appeared in the register between February and April 2016 but who had arrived before January 2016. Finally, in order to have a sufficient number of minors and families in the sample, in tranche M4-2, only minors who appeared in the register by June 2016 were sampled as anchor respondents.

Table 1: Sample Tranches in the IAB-BAMF-SOEP Survey of Refugees

	Sample Tranches					
	M3-1	M3-2	M4-1	M4-2	M5-1	M5-2
Date of Arrival	Jan. 1, 2013-Jan. 31, 2016	Jan. 1, 2013-Jan. 31, 2016	Jan. 1, 2013-Jan. 31, 2016	Jan. 1, 2013-Jan. 31, 2016	Jan. 1, 2013-Jan. 31, 2016	Feb. 1, 2016-Dec. 31, 2016
Date of Register Entry	By Jan. 2016	Feb. 2016-Apr. 2016	By Apr. 2016	By Jun. 2016	May 2016-Jan. 2017	By Jan. 2017
Anchor Respondent	Adult	Adult	Adult	Minor	Adult	Adult

2.2 Sampling and Field Access to Initial Reception Facilities

Sampling asylum seekers and refugees based on their first address upon arrival in Germany typically means using the address of a centralized "initial reception facility" (*Erstaufnahmeeinrichtung*). Yet after just a few weeks, asylum seekers and refugees are often moved to other federal states, counties, and municipalities according to the EASY quota system and assigned housing until their application process is completed. Accommodations differ widely at the local level, ranging from public housing, refugee hostels (for which they receive vouchers), converted gymnasiums, to private apartments.

After crossing the German border, every person who seeks asylum in Germany is sent to a reception facility. Usually, the registration procedures as well as medical care and examinations are carried out there. However, procedures differ across federal states, meaning that the degree of organization and access to the individuals for the purpose of our interviews vary tremendously. What all these facilities have in common is that each person's stay is a maximum of six months and generally a minimum of six weeks (see §47 AsylG). If a person immigrates from a country that is legally categorized as "safe", their stay at the reception facility may be prolonged until deportation. During their stay in these accommodations, refugees are neither allowed to work (see §61 Abs. 1 AsylG) nor are they allowed to rent an apartment in the area (see §3 Abs. 1 AsylbLG).

The comparatively short period of time spent at the reception facility makes it even more challenging to contact and interview potential respondents. We expected that ignoring the high mobility of refugees in these initial housing conditions would lead to high non-contact rates during fieldwork. We therefore established a procedure to ensure that the time between sampling, transferring information to the fieldwork organization, and

contacting the respondent was reduced to just one week. We randomly sampled 11 reception facilities across the county in tranches M3-1, M3-2, M4-1, and M4-2. Even though contact data in reception facilities were immediately passed to the fieldwork, mobility to subsequent housing was so high in many cases that contact in reception facilities often was unsuccessful resulting in a response rate in these specific cases of just 13 percent. As the average length of accommodation in the initial reception facilities dropped considerably with the decreasing numbers of incoming refugees in 2016 to often one or two weeks only, we refrained from implementing the procedure in the later M5-1 and M5-2 tranches.

3. Fieldwork

Interviewing migrants and refugees in particular poses numerous challenges and requires special fieldwork measures to ensure high survey data quality. In the following, we briefly summarize selected aspects of fieldwork design tailored to the population of asylum seekers and refugees.

3.1 Interviewers and Interviewer Training

Given the specifics of the target population as well as the rather unusual interview setting in public housing, the interviewers required special training. A number of measures were undertaken to meet the needs of both respondents and interviewers.

In advance of the fieldwork, a qualitative pretest was implemented (see Brücker et al., 2016), in which social scientists with training in psychology conducted interviews with recently arrived refugees, many of them likely traumatized, who were living in crowded rooms in public housing. The pretest also identified important topics of forced migration and displacement that made it possible to streamline and limit the overall length of the questionnaire. Additionally, sensitive topics were identified and, if appropriate, left out of the survey later.

Additionally, based on the results of the qualitative interviews, training routines and material were developed for the main fieldwork. Interviewers were provided with a comprehensive interviewer handbook. Incentives were also used differently in this population than with other SOEP survey populations. In the pretest, interviewers were informed that monetary incentives had to be deducted from respondents' social benefits. As a result, monetary incentives were not ultimately used. To respond to these findings,

it was suggested that instead of giving money to the adult respondents, small presents could be given as a thank-you for the household's participation. The gifts were given prior to the interview in order to avoid the impression that the gift is payment or even bribery.

Finally, interviewers were recruited according to slightly different criteria to fit the target population. Around a quarter of the interviewers had a migration background themselves. This is a clear advantage, because they are presumably able to be more empathic, especially on sensitive questions dealing with the personal experience of migration.

3.2 Household Response

The household response rate is around 50 percent across all subsamples (see Table 2; for a detailed overview see Kroh et al., 2017). Locating respondents was a major challenge. The high share of respondents whose address changed is arguably due to the shared accommodations in which many lived. It is more difficult to make initial contact with residents of such accommodations and to contact them again later (e.g., letters sometimes get lost in crowded accommodations), or to find them at all when they have moved to another facility or into private housing. However, taking all these aspects into account, the overall response rate is even more striking and reflects a generally high motivation to take part in the interview once respondents have been contacted. Of all sampled persons who could be contacted by an interviewer, 71.5 percent participated in the survey.

Table 2: Household (Non-)Response in the IAB-BAMF-SOEP Survey of Refugees

	Subsamples		
	M3	**M4**	**M5**
Response	47.0 (1,698)	50.6 (1,638)	54.7 (1,519)
Nonresponse	53.0 (1,912)	49.4 (1,600)	45.3 (1,256)
Not locatable/accessible	33.2 (1197)	25.7 (832)	29.1 (808)
Illness or Nursing Care	1.0 (37)	0.7 (24)	1.0 (28)
Language Problems	4.1 (153)	4.9 (158)	3.7 (102)
No time/refusal	11.9 (429)	15.8 (512)	11.3 (313)
Other	2.7 (96)	2.3 (74)	0.2 (5)
Total	100 (3,610)	100 (3,238)	100 (2,775)

Note: Percentages. Number of households in parentheses.

To counteract potential bias due to non-participation of households and individuals, non-response weighting adjustment was applied. The non-response models implemented to generating non-response weights build on a vast literature estimating patterns of household non-response in the general population (Coleman & Fararo, 1992; Groves, Cialdini, & Couper, 1992; Kroh, Kühne, Siegers, & Belcheva, 2018), among migrants (Deding, Fridberg, & Jakobsen, 2008; Kroh, Kühne, Goebel, & Preu, 2015) as well as refugees (Buber-Ennser et al., 2016; Cebulla, Daniel, Devine, & Tipping, 2010; De Maio, Silbert, Jenkinson, & Smart, 2014). The main data sources used to estimate response propensity scores stem from our sampling frame, the Central Register of Foreigners. We made use of the anchor respondent's: 1) asylum status at the time of sampling, 2) country of origin, 3) gender, 4) date of arrival in Germany, and 5) age. In addition to individual-level data, we relied on geographically aggregated data from external databases at the county ("INKAR", BBSR, 2018) and municipality level (Regionaldatenbank Deutschland, Statistische Ämter des Bundes und der Länder, 2018). These data sources contain information on regional socio-economic activity (e.g., property prices, GDP) and population characteristics (e.g., asylum seeker benefits). Finally, interviewers were asked to complete a questionnaire on each household they had attempted to contact. From this, we were able to gain a picture of the household's physical surroundings and the interviewer's feelings about these surroundings for all households of the gross sample.

Logistic regression analysis with cluster-robust standard errors was used to estimate response propensities. Comprehensive documentation on all the variables used is provided in (Kroh et al., 2017). Fortunately, only a few variables systematically explained variance in response behavior, indicating only small differences between respondents and non-respondents. One factor that improved response rates was if the interviewers felt safe when arriving at the accommodations and if they rated the housing as being in "very good" or "superior" condition. Besides these interviewer-related factors, characteristics of the respondents affected response behavior as well. Respondents whose asylum application was still pending had a higher chance of responding to an interview. Respondents living in shared accommodations had a higher rate of non-response.

The final non-response weights were combined (= multiplied) with sample design weights that balance unequal sampling probabilities due to the disproportional sampling

design. This combined-weight was then post-stratified by applying the raking technique (also known as "iterative proportional fitting", Deville, Särndal, & Sautory, 1993) with respect to known marginal distributions derived from the sampling frame. In this regard, the raking process included distributional information on the country of origin (seven groups), gender (two groups), age (fourteen groups), date of arrival (twelve groups) and region (twelve groups). For more details on all steps in the weighting procedure see (Kroh et al., 2017).

3.3 Translation of Survey Instruments and Provision of Audio Files

Besides the challenges of sampling a highly mobile population in a timely manner, it was also necessary to take into account that many respondents would probably not have sufficient language skills to take part in interviews in German. Therefore, all interview materials (letters, flyers, and questionnaires) were provided in seven different languages, including German (see Table 3). For the translation of the materials, two professional translators did the translations for each language. First, a German version of the questionnaire was developed. It was then translated into English. One of those versions (English or German) was then the basis for all further translations. Again, two translators each produced a translation, separately. One of the two created a harmonized version, and this was given to the other, who had the opportunity to comment and correct mistakes.

During the interview, German and the respondent's language were displayed on the screen. Thanks to the CAPI mode (computer assisted personal interviewing), interviewer and respondent were able to look at the screen at the same time. Thus, language barriers were considerably minimized (for further details see Jacobsen, 2018).

Table 3: Use of Visual Translations (left) and Audio-Files (right) in Net Sample

Visual Translation	Percent (absolute)	Audio-Files	Percent (absolute)
German/English	14.1 (956)	With every question	7.5 (506)
German/Arabic	65.8 (4,457)	With around 2/3	5.9 (399)
German/Farsi	14.2 (963)	With around half	4.1 (276)
German/Pashto	1.0 (64)	With fewer than half	8.8 (596)
German/Urdu	1.8 (123)	Not at all	73.8 (5,002)
German/Kurmanji	3.2 (216)		
Total	100 (6,779)	Total	100 (6,779)

Due to the fact that a significant level of illiteracy was anticipated in the population, additional audio files for each language were provided. These audio files were implemented into the CAPI system and were produced and recorded by the same translators who had produced the written translations. Next to each question, scale, or subsequent explanation, there was an icon to click on to listen to the audio file.

Finally, if anything in the interviewing process proved to be problematic, the fieldwork agency also provided a hotline staffed by professional interpreters who could help with initial contact as well as interviews.

Table 3 displays the usage of the written translation in the respective languages. Arabic was used most frequently. This reflects the composition of the target population, of which a large share came from Syria. Farsi, which is spoken in Afghanistan and Iran, was used relatively often as well. It is striking that around 14 percent chose English although there are very few native English speakers in the net sample. Here, we assume that many respondents were relying on their second language. Our results show that most respondents (and interviewers) did not make use of the audio files (74 percent), and only 8 percent used them with every question. At the end of the interview, interviewers were asked to rate the support provided by the tools. The written translation was generally perceived by interviewers to be particularly helpful.

4. Conclusion: Hard-to but not Impossible-to Sample

The IAB-BAMF-SOEP Survey of Refugees aims at filling a gap in research data on the influx of refugees to Europe. The project draws on the Central Register of Foreigners as a sampling frame. Although, access to national register data ensures convenient and controlled means of sampling target population members, the sampling design also had to address a number of challenges. First, there was the issue of gaps in coverage, as the register had been unable to keep up with the migration influx. We addressed this by drawing multiple sub-samples from the register at consecutive points in time. Second, refugees in general, but especially those housed in initial reception facilities are a highly mobile population whose addresses change relatively frequently. A tailored sampling procedure was implemented in order to shorten the time between sampling and initial contact to about a week.

To sum up, for future projects dealing with moving target populations, we recommend sampling in "tranches" and "timely sampling", that is, dramatically reducing the time between sampling and interviewing. However, it should be kept in mind that not only sampling, but sample design weighting is more complex in this case, because the different tranches overlap. In order to ensure representativeness, we had to account for this in our weighting and post-stratification procedures.

Turning to the fieldwork and interviewing of refugees, a number of challenges arose as well. Interviewing refugees in public, centralized housing units was generally more difficult. Interviewers needed to first gain access to the accommodations (often from security guards), find the sample members, and conduct an interview under unusual conditions (for instance, while sitting on a camp bed in a gymnasium). Fortunately, constant monitoring of the fieldwork and feedback from the interviewers themselves showed that the interviewers performed very well in contacting and interviewing respondents, even under harsh conditions.

Furthermore, using the interviewer questionnaire in the analysis of non-response revealed that the assessments given by the interviewers themselves can play a key role in understanding household non-response. Having this information is extremely valuable, especially when interviewing a target population about whom little is known. Using such tools provides useful insights that can help in assessing the quality of the data. The ongoing feedback given throughout the fieldwork phase also provided the research consortium with valuable information. This allowed, for instance, for the change in the incentive strategy mentioned above.

Besides interviewer characteristics, the translation of field instruments was key in surveying a recently immigrated target population. However, during fieldwork, we also learned that an insufficient number of languages had been provided (over 30 percent of respondents had no match for their mother tongue) and that some languages were more useful than others. Therefore, in upcoming projects, a focus of effort should be on the selection of languages to translate in order to avoid wasting resources.

Although our sampling strategy breaks new ground, several limitations should be noted. First, respondents who were supposed to leave the country but went into hiding, who sought sanctuary in churches, or who lived in other forms of informal "protection" were

not included. Furthermore, even though unaccompanied minors are a part of the target population and a particular focus of public and policy interest, they could not be surveyed due to ethical considerations and are therefore not part of the net sample.

The IAB-BAMF-SOEP Survey of Refugees represents an innovative project for surveying a hard-to-reach and hard-to-interview population. It is our hope that this project and the findings discussed in this paper will function as a practical framework and contribute to the survey design of future studies investigating similar populations.

Reference

BAMF. (2015). Asylgeschäftsstatistik für den Monat Dezember 2015. Retrieved from https://www.bamf.de/SharedDocs/Anlagen/DE/Downloads/Infothek/Statistik/Asyl/201512-statistik-anlage-asyl-geschaeftsbericht.pdf?__blob=publicationFile, last visited: 01.10.2018

BAMF. (2018). *Das Bundesamt in Zahllen 2017. Asyl, Migration und Integration*. Nürnberg.

BBSR. (2018). Indikatoren und Karten zur Raum- und Stadtentwicklung in Deutschland und in Europa. Retrieved from http://www.inkar.de/, last visited: 01.10.2018

BMI. (2016). 890.000 Asylsuchende im Jahr 2015. Retrieved from https://www.bmi.bund.de/SharedDocs/pressemitteilungen/DE/2016/09/asylsuchende-2015.html, last visited: 01.10.2018

Brücker, H., Kunert, A., Mangold, U., Kalusche, B., Siegert, M., & Schupp, J. (2016). Geflüchtete Menschen in Deutschland - eine qualitative Befragung. *IAB-Forschungsbericht*(9).

Buber-Ennser, I., Kohlenberger, J., Rengs, B., Zalak, Z. A., Goujon, A., Striessnig, E., . . . Lutz, W. (2016). Human Capital, Values, and Attitudes of Persons seaking Refuge in Austria in 2015. *PLoS One, 11*(9), 1-29.

Cebulla, A., Daniel, M., Devine, C., & Tipping, S. (2010). Survey of New Refugees. Retrieved from http://doc.ukdataservice.ac.uk/doc/6556/mrdoc/pdf/6556_technical_and_data_notes.pdf, last visited: 02.04.2020

Coleman, J. S., & Fararo, T. J. (1992). *Rational Choice Theory. Advocacy and Critique*. Thousand Oaks: Sage.

De Maio, J., Silbert, M., Jenkinson, R., & Smart, D. (2014). Building a New Life in Australia. *Family Matters*(5), 5-14.

Deding, M., Fridberg, T., & Jakobsen, V. (2008). Non-Response in a Survey Among Immigrants in Denmark. *Survey Research Methods, 2*(3), 107-121. doi:10.18148/srm/2008.v2i3.98

Deville, J.-C., Särndal, C.-E., & Sautory, O. (1993). Generalized Raking Procedures in Survey Sampling. *Journal of the American Statistical Association, 88*(423), 1013-1020. doi:10.2307/2290793

Groves, R. M., Cialdini, R. B., & Couper, M. P. (1992). Understanding the Decision to Participate in a Survey. *Public Opinion Quartely, 56*(4), 475-495. doi:10.1086/269338

Jacobsen, J. (2018). Language Barriers During the Fieldwork of the IAB-BAMF-SOEP Survey of Refugees in Germany. In D. Behr (Ed.), *Surveying the Migrant Population: Consideration of Linguistic and Cultural Issues*. Köln: GESIS.

Kroh, M., Kühne, S., Goebel, J., & Preu, F. (2015). The 2013 IAB-SOEP Migration Sample (M1): Sampling Design and Weighting Adjustment. *SOEP Survey Papers*(271).

Kroh, M., Kühne, S., Jacobsen, J., Siegert, M., & Siegers, R. (2017). Sampling, Nonresponse, and Integrated Weighting of the 2016 IAB-BAMF-SOEP Survey of Refugees (M3/M4) - Revised Version. *SOEP Survey Papers*(477).

Kroh, M., Kühne, S., Siegers, R., & Belcheva, V. (2018). SOEP-Core - Documentation of Sample Sizes and Panel Attrition (1984 until 2016). *SOEP Survey Papers*(480).

Lohr, S. L. (2010). *Sampling: Design and Analysis. 2nd edition*. Boston: Brooks/Cole.

Statistische Ämter des Bundes und der Länder. (2018). Regionaldatenbank Deutschland. Retrieved from https://www.regionalstatistik.de/genesis/online/logon, last visited: 01.10.2018

von Gostomski, C. B., & Pupeter, M. (2008). Zufallsbefragung von Ausländern auf Basisdes Ausländerzentralregisters: Erfahrungen bei der Repräsentativbefragung "Ausgewählte Migrantengruppen in Deutschland 2006/2007" (RAM). *Methoden, Daten, Analysen (mda),* 2(2), 149-177.

5 Article 2: If They Don't Understand the Question, They Don't Answer

Preface

In the first paper, we proposed a novel sampling approach to ensure that the IAB-BAMF-SOEP Survey of refugees allows for inference. We further emphasized the importance of translated field instruments to meet the needs of the target population. The second paper will discuss the question of language demands in-depth and analyze how a mismatch between mother tongue and survey language can affect response behavior. One problematic response behavior is item nonresponse.

Although the literature on item nonresponse is vast, little attention has been given to item nonresponse in a multi-linguistic setting. Most research on this matter is concerned with the quality of translation in order to allow comparability across languages. However, in some surveys, not all languages used within the population of interest can be provided due to budget constraints or a diverse target population. That this can lead to unit nonresponse has been shown by Kroh et al. (2017). However, Kroh and colleagues also indicate that some respondents still participated in the survey who could not complete the questionnaire in their mother tongue. How this affects response behavior is analyzed in the next article of this dissertation.

If they don't understand the question, they don't answer. Language mismatch in face-to-face interviews

The article is submitted to and currently under review at *methods, data, analyses.*

Abstract

The provision of translated field instruments is a crucial aspect to ensure data quality in surveys with a multi-linguistic target population such as surveys on recent immigrants. Failure to address this can result in a mismatch between the survey language and the respondent's mother tongue. Respondents who cannot fully understand the content of questions are likely not to give an answer and will be underrepresented in analyses. By using a survey on refugees in Germany, this paper explores the correlation of the absence of the respondents' mother tongue on item nonresponse. Further, this article investigates whether supplementary audio recordings in the same language as the written questions can reduce item nonresponse when the mother tongue is not available.

In order to answer the research questions, all missing answers per individual are counted and analyzed by means of poisson regression analyses. In a second step, the likelihood of item-nonresponse for single items is estimated as well. Results show that a language mismatch as well as the usage of audio recordings increase item nonresponse. It is proposed that this is, first, due to not entirely understanding the questions, and second, due to an increasing respondent's burden.

1. Introduction

The provision of translated field instruments is a crucial aspect to ensure data quality in surveys with a multi-linguistic target population such as surveys on recent immigrants. However, due to budget constraints sometimes not all occurring languages can be provided by field institutes. For some respondents this leads to the absence of their mother tongue as a survey language and they need to choose a second or third language. It was previously shown that answering questions in a second language can lead to a changing response behavior (Elliott, Edwards, Klein, & Heller, 2012; Peytcheva, 2018; Zavala-Rojas, 2018). However, what has been neglected so far is item nonresponse as the result of the absence of the mother tongue. I hypothesize that answering questions in a second or third language hampers the comprehension of survey questions and in turn increases item nonresponse.

I aim to analyze the effect of language mismatch on item nonresponse in the first wave of the German IAB-BAMF-SOEP Survey of Refugees (Kühne, Jacobsen, & Kroh, 2019), a survey that used translated questionnaires and audio recordings[1] to facilitate participation among illiterate respondents. Audio recordings were also expected to facilitate participation among respondents whose mother tongue was not available. The underlying reasoning was that with audio recordings, respondents could rely on listening comprehension, which for some people might be better developed than reading comprehension in language learners. The additional option of listening also offers respondents a second means of understanding a question. Therefore, I am further interested in the question of whether providing audio recordings in addition to written questions helps to decrease item nonresponse in cross-linguistic research.

My results show that language mismatch in general as well as the usage of audio recordings lead to a higher number of unanswered questions.

2. Known Reasons for Item Nonresponse and the Usage of Audio Recordings

[1] Thus, the respondent could switch from CAPI to either CASI or ACASI. A more detailed description is given in section 4.1.

When addressing item nonresponse as a measure to assess data quality, the research to date has generally discussed sensitive questions, cognitive abilities, and respondent burden as possible causes.

Sensitive questions deal with issues that are controversial or private in nature (Coutts & Jann, 2011, p. 177; Jann, Jerke, & Krumpal, 2012). Krumpal reports that questions about income or voting behavior are especially sensitive and subject to high item nonresponse (Krumpal, 2013, p. 2027). Sensitive questions for migrants may deal with sexual or illegal behavior (Pennell, Hibben, Lyberg, Mohler, & Worki, 2017, p. 189). Respondents often skip such questions because they do not want to reveal their true behavior or opinions (Heise, 2010; Krumpal, 2013; Lee, 1993; Lensvel-Mulders, 2008; Nederhof, 1985). Coutts and Jann show that such response behavior can be minimized by applying unmatched count techniques (UCT) (Coutts & Jann, 2011), whereas Jann et al. propose the crosswise model (CM) to reduce social desirable answers (Jann et al., 2012; see also the special issue edited by Jann, Krumpal, & Wolter, 2019). However, especially with regard to cross-cultural research, sensitive topics differ widely and are therefore difficult to identify (Greenfield, 2009; Lyberg et al., 2014, p. 88).

Research on cognitive abilities as a potential cause of item nonresponse has hypothesized that people differ in their ability to answer questions (Converse & Presser, 1986, p. 10). Ahlmark et al. show that unit nonresponse is generally higher among low educated respondents (Ahlmark et al., 2015), a finding which is supported by Mostafa et al. (Mostafa & Wiggins, 2015). Furthermore, Eckman and Kreuter report that cognitively burdensome question formats such as looping questions may lead to item nonresponse (Eckman & Kreuter, 2018). When applying these results to surveys of recent immigrants, questions might need to be simplified due to respondents' potentially limited skills in the local language.

Closely connected to the issue of cognitive abilities is that of respondent burden (see Krosnick, 1991; Krosnick, 1999, p. 546; Tourangeau, Kreuter, & Eckmann, 2015). If the burden (e.g. length of a questionnaire) of an interview is high, respondents tend to skip questions. Kreuter et al., for example, report that this is the case especially for long item batteries that follow questionnaire filters (Kreuter, McCulloch, Presser, & Tourangeau, 2011). I argue that in surveys of recent immigrants, few respondents will actively seek to reduce the burden by avoiding filter questions that would make their interview longer.

In order to do this, they would have to be familiar with questionnaire design in general and filter structures in particular. Recent immigrants (i.e. language learners) might not be familiar with the design of local surveys. Instead, I expect that language mismatch increases survey burden, because it is more demanding to answer questions in a language in which the respondent has not yet developed advanced reading and/or listening skills.

To sum up, the reasons for item nonresponse are manifold and may be culturally dependent. I assume that especially the question of cognitive ability and respondent burden correlate with language mismatch.

2.1 Audio Recordings and Text-to-Speech in a Cross-Linguistic Setting

There are two ways to make a written translation audible. First, there are audio recordings with actual humans reading out aloud and, second, there is text-to-speech technology, which generates an audible version digitally (Couper, Berglund, Kirgis, & Buageila, 2016). From a theoretical point of view, both means should facilitate survey participation because they offer an additional way to understand a questionnaire. However, there is mixed evidence regarding their utility. Kieruj et al. report that text-to-speech technology can be a useful addition to written surveys but that text-to-speech can increase the risk of socially desirable answers (Kieruj, Mulder, Wijnant, Douhou, & Conrad, 2013). Couper et al. (Couper, Tourangeau, & Marvin, 2009) confirm previous research (Tourangeau & Smith, 1996) on audio computer-assisted self-interviewing (ACASI) by showing that respondents do not use audio recordings as much as expected. The advantages of such technologies are therefore small compared to the costs. However, in contrast to Kieruj, they find that audio recordings modestly decrease socially desirable answers. This is in line with earlier research on the subject (Turner et al., 1998). Unfortunately, these studies do not explore whether their findings are robust in cross-cultural and cross-linguistic research. Critically, however, their results suggest that audio recordings, as well as text-to-speech, must be adjusted to the linguistic needs of potential respondents.

In sum, there are few studies to date that have thoroughly investigated the utility of audio recordings. Their findings indicate that the utility is rather limited.

3. Foreign Languages in Surveys Introduce New Causes of Item Nonresponse

A crucial but little-discussed question surrounds the choice of language in surveys. In many cases, when foreigners make up a significant part of the target population, questionnaires are translated (for example Jesske, 2018; Kroh, Kühne, Goebel, & Preu, 2015; Kühne & Kroh, 2017). Here, it is important to ask how researchers ensure that each version conveys the same meaning in the multiple languages used (see for example the Cross Cultural Survey Guidelines CCSG, 2016; Harkness, 2003, 2007; Harkness, Pennell, & Schoua-Glusberg, 2004). In addition to cognitive pretesting (Comanaru & d'Ardenne, 2018, p. 18), researchers increasingly propose the use of expert reviews to validate that translations make sense semantically and incorporate the intended meaning (for example Goerman, Meyers, & Trejo, 2018). However, besides concerns regarding the quality of the translation, there is the question of language mismatch.

Even if questionnaire development can reduce translation bias through pretesting and expert reviews, this does not ensure that all respondents will be provided a version of the questionnaire in their native language. As has been previously shown, answering in a second language can lead to changes in the response behavior compared to answering in the mother tongue (Peytcheva, 2018; Zavala-Rojas, 2018). However, the possibility that respondents do not understand the question and therefore do not answer has been neglected so far.

To the best of my knowledge, a thorough analysis of how language issues are linked to item nonresponse does not yet exist in the literature. Referring to the previous research on survey translation and on item nonresponse I assume, first, that language mismatch increases respondent burden, and second, that it hampers understanding the content of the survey items.

Hypothesis 1: People who cannot use their mother tongue have a higher risk of not responding to a question.

The research partners of the IAB-BAMF-SOEP Survey of Refugees anticipated this problem and therefore provided audio recordings. In addition to facilitating participation among illiterate respondents, it might facilitate participation among those whose native language was not available among the translated versions of the questionnaire. Since this meant that they had an additional version of the questionnaire available to them that they did not have to read, this option was assumed to reduce

respondent burden. Providing the questions in audio as well as written form gives respondents a second possibility to understand a question. However, it is an empirical question whether this actually works when the recording is not in their mother tongue. Assuming that audio recordings reduce respondent burden, I hypothesize:

Hypothesis 2: *When the language of the audio recording matches the respondent's mother tongue, respondents are more likely to answer a question.*

4. The IAB-BAMF-SOEP Survey of Refugees – a Multilingual Survey

I work with the first wave of the IAB-BAMF-SOEP Survey of Refugees. Consisting of 4,465 respondents in 3,289 households[2], it is a household survey of refugees who arrived in Germany between January 2013 and January 2016. The field phase was from June to December 2016.

All field material was provided in seven languages: German, Arabic, English, Farsi/Dari, Kurmanji, Pashto, and Urdu. These are the languages identified as being spoken frequently among the refugee population in Germany. The translation process took place in three steps. All field material was initially developed in German. Thereafter, two translators, working separately, translated the German version into English. The basis for all further translation was either the German or the English version. Again, all versions were developed by two professional translators working separately, whereas one translator produced a unified questionnaire and spoke in the audio recordings (for more details please see also Jacobsen, 2018).

The sampling frame is the Central Register of Foreigners in Germany (AZR) (see Gostomski & Pupeter, 2008). The overall response rate is around 50 percent. The research partners of IAB, BAMF, and SOEP[3] report that around nine percent of total household nonresponse was due to "language problems" (Kühne et al., 2019). They also provide figures indicating that a significant share of the adult net sample are from countries including Somalia, Eritrea, the Balkans, and Russia (Kroh, Kühne, Jacobsen, Siegert, & Siegers, 2017, p. 30), where none of the languages provided are actually spoken. Even though not all languages that are common within the German refugee population

[2] https://doi.org/10.5684/soep.v34

[3] Institute for Employment Research (IAB), Federal Office for Migration and Refugees (BAMF), Socio-economic Panel (SOEP)

were part of the translation process, a significant share of respondents whose mother tongue was not provided participated in the survey (this is indicated by Kroh et al., 2017 who provide numbers on the country of origin).

A more detailed overview on sampling, the target population, and calculation of response rates is given by Kroh et al. and Kühne et al. (Kroh et al., 2017; Kühne et al., 2019).

4.1 The Application of Translated Field Instruments

The mode of the survey was CAPI, CASI, or ACASI[4]. Prior to the interview, the respondent had to decide which additional language (besides German) was needed and choose from the translations provided. Respondents were required to decide which translated version to use and could not change this decision in the course of the interview.

The chosen translation and the German version were presented side by side on the computer screen. Usually the interviewer would ask a question and record the answer, thus CAPI was the default. In the event that the interviewer did not speak the respondent's language and the respondent did not speak German, the CASI replaced the CAPI mode and the respondent could fill out the questionnaire himself. Further, if the written questions, for any reason, could not be used by the respondent, they could switch to ACASI. Because audio recordings were provided for each question and subsequent explanation separately, the respondent could easily mix CASI and ACASI. Respondents simply had to click a button and the question was read out aloud in the chosen language (for a detailed overview and a picture of the CAPI/CASI screen see Glemser, Huber, & Leven, 2017, p. 51). In this event, the interviewer did not leave, but rather gave general guidance instead of reading all questions aloud. Producing audio recordings of the answers was not possible.

4.2 Usage of Translation and Audio Recordings

The extent to which different translations were chosen varies widely. As Table 1 shows, most people chose Arabic, followed by English.

[4] Computer-assisted personal-interview (CAPI), computer-assisted self-interview (CASI), audio computer-assisted self-interview (ACASI).

Table 1: Choice of Secondary Interview Language

Translation	Frequency (percent)
German/English	727 (16.3)
German/Standard Arabic	2,891 (64.8)
German/Farsi	571 (12.8)
German/Pashto	47 (1.1)
German/Urdu	77 (1.7)
German/Kurmanji	152 (3.4)
Total	*4,465 (100)*

Source: IAB-BAMF-SOEP Survey of Refugees, author's calculations

As not all occurring languages could be provided, I conclude that many chose English because of the absence of their mother tongue. In general, about 85 percent used the written translation; 58 percent used them for every question (Table 2). Overall, 26 percent used the audio recordings, but not necessarily for every question; 6 percent used the audio recordings for every question. Cross-tabulating the usage of the translation and the usage of audio recordings indicates that 77 percent who used the audio recordings with every question also used the visual translation with every question (not displayed as a table).

Table 2: Frequency of Chosen Translation Tool Based on Interviewer Observation

	Written	Audio
	Frequency (percent)	
With every question	2,583 (57.9)	282 (6.3)
With two-thirds	532 (11.9)	280 (6.3)
With half	350 (7.8)	180 (4.0)
With less than half	343 (7.7)	411 (9.2)
Not at all	657 (14.7)	3,312 (74.2)
Total	*4,465 (100)*	*4,465 (100)*

Source: IAB-BAMF-SOEP Survey of Refugees, author's calculations

As Table 3 indicates, around 40 percent of the respondents could not find a match for their mother tongue. In particular, Arabic, Farsi, Pashto, and Kurmanji fit the population well in relative terms, whereas English seems to be spoken less as a mother tongue. Examining the reported mother tongue of respondents (Table A1 in the appendix), we see that particularly respondents from Eritrea, Somalia, and the Balkans were unable to find their mother tongue, languages such as Tigrinya, Albanian, and Somali.

Table 3: Match of Mother Tongue and Chosen Translation

	Match	
Translation	No (percent)	Yes (percent)
English	716 (98.5)	11 (1.5)
Arabic	898 (31.1)	1,993 (68.9)
Farsi	54 (9.5)	517 (90.5)
Pashto	8 (17.0)	39 (83.0)
Urdu	35 (45.5)	42 (54.6)
Kurmanji	20 (13.2)	132 (86.8)
Total	*1,731 (38.8)*	*2,734 (61.2)*

Source: IAB-BAMF-SOEP Survey of Refugees, author's calculations

4.3 Operationalization

To test whether a language mismatch leads to item nonresponse, I chose to count the missing values for each person in the survey. For this procedure, I only used variables that were not dependent on looping of filter questions in order to avoid nonresponse that might be due to the questionnaire design. Further, I did not count "don't know" as missing values when it was displayed as a valid answer category, and also did not include interviews that were not completed. Unfortunately, I cannot identify whether some item nonresponse is due to questionnaire routing errors, data processing, or triggered by interviewers who wanted to decrease their workload. In total, I counted nonresponse for 170 variables. The number of missing values of the respondents ranges from zero to 125

(mean = 7.9, SD = 14.3). Therefore, I apply Zero Inflated Poisson Regression analysis (ZIP) (Lambert, 1992) with the count of unanswered questions ($y_i \sim Poi(\mu_i), i \in \{1, \dots, n\}$) as the dependent variable. This is necessary because the count variable has a small mean value compared to the maximum and a modal value of zero (for a histogram see Figure A1 in the appendix). In a ZIP regression the probability of an individual i having j counts is assumed to be

$$Pr(y_i = j \mid t_i, x_{1i}, \dots, x_{ki}) = \begin{cases} \pi_i + (1-\pi_i)\, exp(-\mu_i) & if\ j = 0 \\ (1-\pi_i)\, \frac{exp\,(-\mu_i)}{y_i!} \cdot \mu_i^{\,j} & if\ j > 0 \end{cases} \quad (1)$$

with π_i being the probability of individual i to have no nonresponse, and μ_i being the Poisson component

$$\mu_i \mid t_i, x_{1i}, \dots, x_{ki} = exp(ln(t_i) + \beta_1 x_{1i} + \cdots + \beta_k x_{ki})) \quad (2)$$

with t_i being the number of questions that could have been answered, $x_{1i}, \dots, x_{ki}$ the regressor variables and $\beta_1, \dots, \beta_k$ the coefficients of the covariates, respectively.

To understand the reasons for item nonresponse, looking at the total individual amount of unanswered questions by an interviewee only tells half the story. Further, it is helpful to assess the determinants for item nonresponse regarding variables that have an exceptionally high share of missing values. To deepen my analysis, in a second step I evaluate the event of a question l to remain unanswered ($z_l \in \{0,1\}$) where 1 denotes that the question is not answered. Therefore, a linear probability regression model with a random effect for the interviewer (j) and a random effect for the respondent (i) is applied[5]:

$$\Pr(\bar{Y}_l) = \varsigma_j + u_i + x_{1l}\beta_1 + \cdots + x_{kl}\beta_k + \varepsilon_l \quad (3)$$

[5] Applying logistic regression analysis with two random effects and the amount of control variables used here (see next paragraph) is not feasible for Stata due to model complexity. Reducing complexity would have resulted in discarding interviewers that conducted fewer than 50 interviews (around 2/3 of all interviewers). Although, a linear probability model has some short comings in predicting binary outcomes, it also has several advantages, especially with complex model structure. Moreover, Mood shows that linear probability models produce similar outcomes in comparison to classical logistic regression analysis (Mood, 2010). However, in order to make sure that the choice of estimation does not produce an artefact I also estimated the less complex logistic model. The conclusions are robust to different model specifications (not displayed as a table).

Where $x_{1l}, \ldots, x_{kl}$ denote the predictor variables. The parameters $\beta_0, \ldots, \beta_k$ are the fixed regression coefficients and have to be estimated. ς_j denotes the random intercept for the interviewer and u_i the random intercept for the participants and ε_l the residual. Nesting the questions not only in respondents but also in interviewers is important to account for interviewer effects (Kosyakova, Olbrich, Sakshaug, & Schwanhäuser, 2019). In order to have sufficient variation on the interviewer level, I exclude those interviewers that conducted less than 10 interviews (resulting in a dismissal of 83 observations). Additionally, I only use variables that have more than 10 percent missing values and are not affected by filter questions. In total, I identified 26 different questions that each have between 10.2 to 27.1 percent missing values (see Table A2 in the appendix for topics, the exact wording, scales, and share of missing values (see also Jacobsen, Klikar, & Schupp, 2017)). In the end, I included 26 questions that could have been answered by a total of 4,382 respondents (not accounting for missing values of independent variables). Therefore, 113,932 distinct observations which are nested in 4,382 respondents, nested in 79 interviewers are used.

Key explanatory variables for all models are, first, a binary variable indicating whether the provided translation matched the mother tongue and, second, if the audio recordings were used. This information is given by the interviewer, but only as a general estimation and not for every question (see again Table 2). Therefore, for the multivariate analyses, I differentiated between "used with every question" and any category that indicates reduced usage. With this approach, I ensure that audio recordings were actually used with the questions analyzed. I also insert an interaction term of the two variables to test whether the audio recordings are only useful when they match the mother tongue.

4.4 Control Variables

In order to control for different abilities and cognitive potentials of the respondent, I used self-reported German skills, educational level (ISCED; (Jaworski, Pagel, & Schupp, 2017), and an estimation of functional illiteracy (Rother, Schacht, & Scheible, 2017). In order to control for a possible respondent burden, I measured the length of the interview and the type of housing (group vs. private). This is important because group housing for refugees is often crowded, making it likely that an interview is interrupted frequently. I assume this lengthens the interview and thus increases the burden.

With regard to sensitivity, I controlled for religion and mental distress (PHQ-4 (Kroenke, Spitzer, Williams, & Löwe, 2009; Perez & Baffour, 2018)). This is necessary because some variables have a religious and psychological connotation. I assume that deeply religious and mentally unstable respondents experience anxiety or distress when talking about such subjects. For the same reason, I controlled for the presence of bystanders during the interview because the respondent might be intimidated or simply not want to talk about sensitive subjects in front of others. For cultural particularities, I controlled for the perceived level of democracy in the home country (e.g. freedom of speech). This is a proxy for different norms and values that might affect response behavior in general, such as responses to sensitive topics. I further controlled for the gender and whether the gender of the respondent matches the gender of the interviewer. This is important in order to rule out that aspects of response behavior are due to gender sensitivity (Benstead, 2013). I also controlled for residence status, because people without refugee status (e.g. asylum seekers or people with suspension of deportation) might be more affected by social desirability. Finally, I controlled for age. For a more detailed description and frequencies of the variables, see Table A3 and Table A4 in the appendix.

For the second, linear probability, model, I additionally controlled for the topics of the unanswered questions, in order to identify whether the effect is confounded by especially sensitive topics. The groups comprise “Month of Immigration”, “Locus of Control”, “Political Situation in Country of Origin”, “Forms of Government”, “Attitudes towards Democracy”, “Attitudes towards Gender Equality” (see Table A2 in the Appendix for more details). Moreover, I exclude the variable controlling for the level of democracy in the home country so that independent and dependent variables are not the same.

5. Results

In order to maintain clarity in this section, all models are presented without showing the effect sizes of control variables. Complete models are provided in the Appendix. Table 4 displays the Poisson regression estimates of the Zero Inflated Poisson Regression analysis with the count of unanswered questions as the dependent variable. It indicates that having a match with the mother tongue decreases the number of not answered questions by a respondent (effect size = -0.30). This effect size means that, holding all other

variables constant, having a match with one's mother tongue would decrease the expected number of unanswered questions by a factor of 0.74.

Table 4: Zero Inflated Poisson Regression Analysis

	Zero Inflated Poisson Regression (ZIP) on number of unanswered Questions [a] [b] (unstandardized coefficients) SE in parenthesis			
Independent Variables	Model 1a		Model 1b	
Match of mother tongue, ref = no	-.30***	(.02)	-.31***	(.02)
Audio recording used, ref = not used	.14***	(.01)	.08	(.05)
Interaction Audio recording*Match	-	-	.08	(.06)
Chi^2	1472		1473	

Source: IAB-BAMF-SOEP Survey of Refugees, author's calculations
***p<0.01 **p<0.05 *p<0.1 (two-tailed-tests)
N (all models) = 3,630, due to missing values, some observations are excluded
Nonzero Observations = 2,505
[a] Control Variables: Gender Match between Interviewer and Interviewee, ISCED, German Language Abilities, Religion, Age, Gender, Duration of Interview, Mental Distress, Residency Status, Type of Housing, Level of Democracy in Home Country, Bystander during Interview, Functional Analphabetism
[b] The estimates of the Poisson part of the function are presented only. The estimation of having no nonresponse is available upon request.
See Table A5 in the Appendix for effects of control variables.

As a robustness check I also inserted an interaction term between the match variable and the chosen language for the interview. Table A6 in the appendix shows that the effect of the language mismatch basically remains unchanged. This means that there is no particular language driving the correlation.

The model analyzing nonresponse of single variables (Table 5) supports the finding of the previous estimation. Questions that were presented to respondents with a language match have a substantial lower probability of not being answered (AME = -0.05). Considering that the assessed variables have between 10 and 27 percent missing values, this effect size is quite substantial.

Table 5: Linear Probability Model on Item Nonresponse with Questions Nested in Respondents Nested in Interviewers

	Linear Probability Regression Analyses with Random Intercepts [a] (Average Marginal Effect, 1 = Nonresponse) SE in parenthesis			
Independent Variables	Model 1a		Model 1b	
Match of mother tongue, ref = no	-0.05***	(0.01)	-0.05***	(0.01)
Audio recording used, ref = not used	0.05***	(0.01)	0.01	(0.02)
Interaction Audio recording*Match	-	-	0.06**	(0.03)
Topic, ref = Month of Immigration				
Locus of Control	-0.06***	(0.00)	-0.06***	(0.00)
Political Situation in CoO	-0.07***	(0.00)	-0.07***	(0.00)
Forms of Government	0.03***	(0.00)	0.03***	(0.00)
Attitudes towards Democracy	-0.03*	(0.00)	-0.03**	(0.00)
Attitudes towards Gender Equality	-0.08***	(0.01)	-0.08***	(0.01)

Source: IAB-BAMF-SOEP Survey of Refugees, author's calculations
***p<0.01 **p<0.05 *p<0.1 (two-tailed-tests)
N = 97,266
N(Respondents) = 3,741, due to missing values, some observations are excluded
N(Interviewer) = 79
[a] Control Variables: Gender Match between Interviewer and Interviewee, ISCED, German Language Abilities, Religion, Age, Gender, Duration of Interview, Mental Distress, Residency Status, Type of Housing, Bystander during Interview, Functional Analphabetism.
See Table A7 in the Appendix for effects of control variables.

In order to show how this result can affect future analyses, I additionally estimated mean predicted probabilities for questions to remain unanswered based on model 1a in Table 5 (without interaction term) with respect to the country of origin of the respondents to which the questions were presented. This is valuable because the country of origin, in many analyses of integration research, serves as a proxy and control variable for effects researchers cannot directly measure; for example, schooling systems or historic experiences like civil wars, economic crises, or colonialism. Therefore, when there is indication that the response probability depends on the country of origin, then the missing values of many dependent variables are most likely not completely at random after controlling for the country of origin. Indeed, the results indicate that the likelihood of item nonresponse increases when a question is answered by respondents from parts of the world where none of the languages provided in the translations are common (see Table 6).

Table 6: Mean Probability for Questions with More Than 10 Percent Missing Values to Remain Un-answered by People from Different Countries of Origin

Country of Origin	Probability in Percent [95 Percent Confidence Interval]
Syria	9 [9.1-9.2]
Iran	10 [9.9-10.4]
Pakistan	11 [10.9-11.4]
Afghanistan	13 [12.6-12.8]
Iraq	13 [12.6-12.9]
Russia	13 [12.4.-12.8]
Somalia	15 [15.1-15.7]
Ukraine	16 [15.6-16.5]
Albania	16 [15.3-15.9]
Kosovo	16 [15.3-15.8]
Macedonia	18 [17.9-18.9]
Eritrea	18 [17.4-17.7]
Serbia	18 [17.5-18.1]
Nigeria	20 [19.5-20.6]

Source: IAB-BAMF-SOEP Survey of Refugees, author's calculations for countries with N >= 20 in IAB-BAMF-SOEP Survey of Refugees based on model 1a in Table 5. Note: As linear probability models can produce estimated probabilities outside 0,1, all estimates above and under are trimmed to 0 and 1 respectively (for 3% of all estimates).

For instance, questions answered by people from Nigeria (who has several different official languages and dialects) have the highest mean nonresponse probability. Questions answered by respondents from Serbia (Serbian) and Eritrea (Tigrinya) have high mean probabilities of remaining unanswered as well.

Estimated mean probabilities regarding the respondents' mother tongue (Table A8 in the appendix) present a similar picture. Questions that were answered by those who speak Somali or Tigrinya have an especially high mean nonresponse probability. As Somali and Tigrinya are common among Somalian and Eritrean respondents, the results support the

effect of a language mismatch. Moreover, the table clearly indicates that those languages who were provided through the questionnaire all produce significantly less nonresponse.

With regard to the second hypothesis on whether audio recordings decrease item nonresponse, I again consider Table 4 first. The model without the interaction term between audio file usage and a language match indicates that the usage of audio recordings mildly increases the number of missing values (effect size of 0.13 = factor of 1.1). After including the interaction term however, the estimate becomes insignificant. Moreover, the interaction term itself is insignificant as well. Because both point estimates are positive however, this could be a hint that audio recordings only increase item nonresponse when they match the mother tongue. But, these results are clearly inconclusive.

As the finding that audio recordings increase item nonresponse is somewhat surprising, as a robustness check I also tested a variable reflecting the usage of audio recordings in more detail instead of a dummy variable (see Table 2 for a more detailed coding). However, the interpretation that the usage of audio recordings increases the number of unanswered questions remains unchanged (for regression results see Table A9 in the appendix).

The linear probability regression analysis in Table 5 supports the finding of the Zero Inflated Poisson Regression analysis. Questions which were answered using audio-recordings are five percent more likely to remain unanswered. However, when the interaction term is included it becomes clear that the effect holds only for those cases in which the audio-record matches the mother tongue (AME = 0.06). This means that the usage of audio recordings increases the risk for questions not to be answered when the recording matches the mother tongue. The size of the effect further indicates that this difference is in fact substantial. However, as 71 percent of those who used the recordings with every question had a match with their mother tongue (not displayed as a table), I assume that this interaction effect is an artefact triggered by self-selection of the respondents. Respondents who did not have a language match probably did not consider the recordings helpful in general and refrained from using them whereas for those with a match, only those used the recordings who have trouble to comprehend the question anyways.

Turning to the topics of the unanswered questions, both models of the linear probability regression reveal that some subjects are more prone to produce item nonresponse than others, for example the question regarding the month of immigration. As this, if several years ago, can be hard to remember, I assume that this reflects issues of retrieval. Therefore, I additionally conclude that items which cover topics that deal with the past are additionally more prone to item nonresponse.

5.1 Discussion: Does a Language Mismatch Lead to Item Nonresponse?

All analyses point to the conclusion that a language mismatch leads to a significant increase in item nonresponse. First, my results indicate that those who have a match for their mother tongue answer more questions in total. Second, the analyses of the nonresponse on single questions also clearly indicate that a mismatch increases the probability of questions to remain unanswered. The effect proved to be independent of the content of the question, meaning that a language mismatch causes item nonresponse in general.

I assume that this is for two main reasons: First, individuals that are proficient in the chosen language have the ability to answer a wide array of questions, including complex ones. Respondents who cannot answer in their mother tongue have a higher barrier to comprehend the content of a question.

Second, I assume that answering in a second or third language requires more effort during the interview. Although respondents with a language mismatch answer fewer questions, they take the same amount of time to finish the interview (mismatch: median = 83 min, IQR[6] = 54 min; match: median = 84 min, IQR = 48 min). Thus, a mismatch appears to increase the burden.

In sum, a language mismatch, first, prevents respondents from understanding a question entirely and, second, increases respondent's burden. As a result, respondents with a language mismatch increase item nonresponse in surveys.

5.2 Do Audio Recordings Facilitate Survey Participation?

Although language mismatch in surveys appears to increase item nonresponse, the effect of audio recordings designed to bridge language gaps is not straightforward. With regard

[6] Interquartile range

to the total answers given by a respondent, my results are somewhat inconclusive. However, the analyses of single questions indicate that the usage of audio recordings increases the risk that questions are not answered when the audio recording matches the mother tongue. However, I assume that this effect is due to self-selection: respondents who do not have a language match do not use the audio recordings in general as they must appear useless. Those, however, who have a match will only use the recordings if they have a problem of comprehension anyway. Therefore, the effect is likely to reflect these issues than the pure effect of the recordings.

Additionally, it is likely that those who use audio recordings first read the question and then listen to it. Therefore, they make twice the effort. Audio recordings clearly lengthen the survey and might lead to a desire to skip questions. Again, this is indicated by the mean interview duration. While the median for respondents who used the audio recordings with every question is 110 minutes (IQR = 62 min) to finish the interview, for others the median is 82 minutes (IQR = 49 min), only. Further, it is also possible that interviewer behavior plays a role. If interviews tend to take longer than usual, interviewers might tend to sway respondents to skip some questions (Kosyakova, Skopek, & Eckman, 2015).

5.3 How Should Mismatch be Addressed in Future Research?

With these findings in mind, I conclude that a mismatch between the mother tongue and survey language increases the likelihood of item nonresponse. Although the paper at hand used a survey with refugees as the target group, a language mismatch can also occur in surveys dealing with other forms of migration or in mono-lingual surveys in general. With regard to the limited budget of such scientific projects, it is impossible to provide all of the languages that occur in a population. Researchers need to invent strategies to deal with this problem *ex ante*. For example, it is important to assess carefully which languages need to be provided. In many cases, proxy information is identifiable through administrative data such as the Central Register of Foreigners in Germany (AZR).

My analysis suggests that if researchers encounter nonresponse in surveys in which respondents' mother tongue is not available, nonresponse reflects these language issues to some extent, and missing values are therefore not random. Thus, solutions to minimize possible biases *ex post* are important. I propose the following solution for

discussion: Applying multiple imputation by chained equations. Due to the fact that the net sample of the IAB-BAMF-SOEP Survey of Refugees provides a vast amount of information on its respondents, it is possible to estimate models that should be able to estimate who answers certain questions and who does not. By applying multiple imputation the probabilistic characteristic of the sample is furthermore maintained.

5.4 Limitations

Some limitations of this study should be noted. Arguably, I underestimate the effect of a language mismatch. First, respondents who declined to participate in the survey due to language problems are not part of this study. If these people had taken part in the survey, the correlation between language mismatch and item nonresponse could have been even greater. Second, some control variables in the multivariate models also have missing values. Bivariate correlations indicate that in some cases, they are not completely random in regard to the missing values of the tested variables. Therefore, the effects are presumably underestimated and even stronger than presented here – making the conclusion of this article even more valuable.

Additionally, I cannot assess with the information at hand whether the dialect of the audio recording matches the specific dialect of the respondent. Especially with regard to the Arabic language, it is crucial to analyze whether the respondents are able to understand Standard Arabic even if their mother tongue is one of the Arabic dialects. This aspect could also increase item nonresponse because it may be that although languages match, dialects do not.

What is more, it is obvious that the variables in the survey have high and, in other cases, low nonresponse. Therefore, although I show that a language match is important, there seem to be variables that are not affected. Why this is the case, even though a language mismatch increases the number of unanswered questions in general, requires further research. At first glance, I would argue that the complexity of the question as well as the perceived sensitivity is crucial. However, question complexity in an extremely diverse target population needs further discussion and research.

Finally, interviewer effects were left largely unmentioned in the discussion of the results. This is because there is little information on them. However, by controlling for whether

the gender of respondent and interviewer is the same and by applying additional multilevel models, I at least controlled for interviewer effects to some extent.

Despite the fact that this study faced some obstacles and therefore has limitations, the findings of increased nonresponse with language mismatch and audio recordings provide new perspectives for multi-linguistic research in the field of survey methodology and new insights into the use of questionnaire translations and audio recordings.

References

Ahlmark, N., Algren, M. H., Holmberg, T., Norredam, M. L., Nielsen, S. S., Blom, A. B., . . . Juel, K. (2015). Survey nonresponse among ethnic minorities in a national health survey – a mixed-method study of participation, barriers, and potentials. *Ethnicity & Health, 20*(6), 611-632. doi:10.1080/13557858.2014.979768

Benstead, L. J. (2013). Effects of interviewer–respondent gender interaction on attitudes toward women and politics: Findings from morocco *International Journal of Public Opinion Research, 26*(3), 369-282. doi:10.1093/ijpor/edt024

CCSG. (2016). Translation: Overview. *Cross-Cultural Survey Guidelines* Retrieved from https://ccsg.isr.umich.edu/index.php/chapters/translation-chapter/translation-overview, last visited: 02.04.2020

Comanaru, R., & d'Ardenne, J. (2018). The development of a research programme to translate and test the personal well-being questions in sylheti and urdu. In D. Behr (Ed.), *Surveying the Migrant Population: Consideration of Linguistic and Cultural Issues*. Köln: GESIS.

Converse, J. M., & Presser, S. (1986). *Survey questions: Handcrafting the standardized questionnnaire*. Newbury Park: Sage.

Couper, M. P., Berglund, P., Kirgis, N., & Buageila, S. (2016). Using text-to-speech (TTS) for audio computer-assisted self-interviewing (ACASI). *Field Methods, 28*(2), 95-111. doi:10.1177/1525822X14562350

Couper, M. P., Tourangeau, R., & Marvin, T. (2009). Taking the audio out of audio-CASI. *Public Opinion Quartely, 73*(2), 281-303. doi:10.1093/poq/nfp025

Coutts, E., & Jann, B. (2011). Sensitive questions in online surveys: Experimental results for the randomized response technique (RRT) and the unmatched count technique (UCT). *Sociological Methods & Research, 40*(1), 169-193. doi:10.1177/0049124110390768

Eckman, S., & Kreuter, F. (2018). Misreporting to Looping Questions in Surveys: Recall, Motivation and Burden *Survey Research Methods, 12*(1), 59-74. doi:10.18148/srm/2018.v12i1.7168

Elliott, M. N., Edwards, W. S., Klein, D. J., & Heller, A. (2012). Differences by survey language and mode among chinese respondents to a CAHPS health plan survey. *Public Opinion Quartely, 76*(2), 238-264. doi:10.1093/poq/nfs020

Glemser, A., Huber, S., & Leven, I. (2017). Report 2016 from kantar public (Samples M1-4). In J. Britzke & J. Schupp (Eds.), *SOEP Wave Report 2016* (pp. 43-54). Berlin: SOEP — The German Socio-Economic Panel study at DIW Berlin.

Goerman, P., Meyers, M., & Trejo, Y. G. (2018). The place of expert review in translation and Questionnaire evaluation for hard-to-count populations in national surveys. In D. Behr (Ed.), *Surveying the Migrant Population: Consideration of Linguistic and Cultural Issues*. Köln: GESIS.

Gostomski, C. B. v., & Pupeter, M. (2008). Zufallsbefragung von Ausländern auf Basis des Ausländerzentralregisters: Erfahrungen bei der Repräsentativbefragung "Ausgewählte Migrantengruppen in Deutschland 2006/ 2007" (RAM). *Methoden, Daten, Analysen, 2*(2), 149-177. doi:http://nbn-resolving.de/urn:nbn:de:0168-ssoar-126673

Greenfield, M. (2009). Reaching gypsies and travellers. *Primary Health Care, 19*(8), 26-27.

Harkness, J. A. (2003). Questionnaire Translation. In J. A. Harkness, F. J. R. Van de Vijer, & P. P. Mohler (Eds.), *Cross-Cultural Survey Methods* (pp. 35-56). Hoboken: John Wiley & Sons.

Harkness, J. A. (2007). Improving the comparability of translations. In R. Jowell, C. Roberts, R. Fitzgerald, & G. Eva (Eds.), *Measuring Attitudes Cross-Nationally: Lessons from the European Social Survey* (pp. 79-94). London: Sage.

Harkness, J. A., Pennell, B.-E., & Schoua-Glusberg, A. (2004). Survey questionnaire translation and assessment. In S. Presser, J. M. Rothgeb, M. P. Couper, J. T. Lessler, E. Martin, J.

Martin, & E. Singer (Eds.), *Methods for Testing and Evaluating Survey Questionnaires* (pp. 453-474). New Jersey: Wiley & Sons.
Heise, D. R. (2010). *Surveying cultures: discovering shared conceptions and sentiments*. Hoboken: Wiley & Sons.
Jacobsen, J. (2018). Language barriers during the fieldwork of the IAB-BAMF-SOEP Survey of Refugees in Germany. In D. Behr (Ed.), *Surveying the Migrant Population: Consideration of Linguistic and Cultural Issues* (pp. 75-84). Köln: Gesis.
Jacobsen, J., Klikar, J., & Schupp, J. (2017). Scales manual IAB-BAMF-SOEP Survey of Refugees in Germany - revised version. *SOEP Survey Papers: Series C - Data Documentation, 475*.
Jann, B., Jerke, J., & Krumpal, I. (2012). Asking sensitive questions using the crosswise model: An experimental survey measuring plagiarism. *Public Opinion Quarterly, 76*(1), 32-49. doi:10.1093/poq/nfr036
Jann, B., Krumpal, I., & Wolter, F. (2019). Editorial: Social desirability bias in surveys - collecting and analyzing sensitive data. *methods, data, analyses, 13*(1), 3-6.
Jaworski, J., Pagel, L., & Schupp, J. (2017). Zuordnung und Überführung erworbener Bildungsabschlüsse in das internationale Bildungsklassifikationssystem (ISCED). *Politikeratung kompakt, 123*, 21-28. doi:http://hdl.handle.net/10419/176772
Jesske, B. (2018). Surveying migrants in the context of the low-income panel PASS. In D. Behr (Ed.), *Surveying the Migrant Population: Consideration of Linguistic and Cultural Issues* (pp. 85-94). Köln: GESIS.
Kieruj, N. D., Mulder, J., Wijnant, A., Douhou, S., & Conrad, F. (2013). *Hearing voices: supporting online questionnaires with text-to-speech technology*. Paper presented at the Priority Programme on Survey Methodology, Bremen.
Kosyakova, Y., Olbrich, L., Sakshaug, J. W., & Schwanhäuser, S. (2019). Identification of interviewer falsification in the IAB-BAMF-SOEP Survey of Refugees in Germany. *FDZ Methodenreport, 2019*(2). doi:10.5164/IAB.FDZM.1902.en.v1
Kosyakova, Y., Skopek, J., & Eckman, S. (2015). Do interviewers manipulate responses to filter questions? evidence from a multilevel approach. *International Journal of Public Opinion Research, 27*(3), 417-431. doi:10.1093/ijpor/edu027
Kreuter, F., McCulloch, S., Presser, S., & Tourangeau, R. (2011). The effect of asking filter questions in interleafed versus grouped format. *Sociological Methods & Research, 40*(1), 88-104. doi:10.1177/0049124110392342
Kroenke, K., Spitzer, R. L., Williams, J. B. W., & Löwe, B. (2009). An ultra brief screening scale for anxiety and depression: The PHQ-4. *Psychosomatics, 50*(6), 613-621. doi:10.1016/S0033-3182(09)70864-3
Kroh, M., Kühne, S., Goebel, J., & Preu, F. (2015). The 2013 IAB-SOEP Migration Sample (M1): Sampling design and weighting adjustment *SOEP Survey Papers: Series C - Data Documentation*(271).
Kroh, M., Kühne, S., Jacobsen, J., Siegert, M., & Siegers, R. (2017). Sampling, nonresponse, and integrated weighting of the 2016 IAB-BAMF-SOEP Survey of Refugees (M3/M4). *SOEP Survey Papers: Series C - Data Documentation*(477).
Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of aittude measures in surveys. *Applied Cognitive Psychology, 5*, 213-236. doi:10.1002/acp.2350050305
Krosnick, J. A. (1999). Survey research. *Annual Review of Psychology, 50*, 537-567. doi:10.1146/annurev.psych.50.1.537
Krumpal, I. (2013). Determinants of social desirability bias in sensitive surveys: a literature review. *Quality & Quantity, 47*(4), 2025-2047. doi:10.1007/s11135-011-9640-9
Kühne, S., Jacobsen, J., & Kroh, M. (2019). Sampling in Times of High Immigration: The Survey Process of the IAB-BAMF-SOEP Survey of Refugees. *Survey Methods: Insights from the Field*. doi:10.13094/SMIF-2019-00005

Kühne, S., & Kroh, M. (2017). The 2015 IAB-SOEP Migration Study M2: Sampling design, nonresponse, and weighting adjustment *SOEP Survey Papers: Series C - Data Documentation*(473).
Lambert, D. (1992). Zero-inflated pisson regression. With an application to defects in manufacturing. *Technometrics, 34*(1), 1-14. doi:10.2307/1269547
Lee, R. M. (1993). *Doing research on sensitive topics*. London: Sage.
Lensvel-Mulders, G. (2008). Surveying sensitive topics. In E. D. d. Leeuw, J. J. Hox, & D. A. Dillman (Eds.), *Internatinoal Handbook of Survey Methodology* (pp. 461-478). New York: Taylor & Francis.
Lyberg, L., Stange, M., Harkness, J., Mohler, P., Pennell, B.-E., & Japec, L. (2014). A review of quality issues associated with studying hard-to-survey populations. In R. Tourangeau, B. Edwards, T. P. Johnson, K. M. Wolter, & N. Bates (Eds.), *Hard-to-Survey Populations* (pp. 82-107). Cambridge: Cambridge University Press.
Mood, Carina. (2010). Logistic Regression: Why We Cannot Do What We Think We Can Do, and What We Can Do About It. *European Sociological Review, 26(1),* 67-82, doi:10.1093/esr/jcp006
Mostafa, T., & Wiggins, R. D. (2015). The impact of attrition and non-response in birth cohort studies: a need to incorporate missingness strategies. *Longitudinal and Life Course Studies, 6*(2), 131-146. doi:10.14301/llcs.v6i2.312
Nederhof, A. J. (1985). Methods of Coping with Social Desirability Bias: A Review. *European Journal of Social Psychology, 15*(3), 263-280. doi:10.1002/ejsp.2420150303
Pennell, B.-E., Hibben, K. C., Lyberg, L. E., Mohler, P. P., & Worki, G. (2017). A total survey error perspective on surveys in multinational, multiregional, and multicultural contexts. In P. P. Biemer, E. d. Leeuw, S. Eckman, B. Edwards, F. Kreuter, L. E. Lyberg, N. C. Tucker, & B. T. West (Eds.), *Total Survey Error in Practice*. Hoboken: Wiley & Sons
Perez, P., & Baffour, B. (2018). Respondent mental health, mental disorders and survey interview outcomes. *Survey Research Methods, 12*(2), 161-176. doi:10.18148/srm/2018.v12i2.7225
Peytcheva, E. (2018). Can the language of survey administration influence respondents' answers? In T. P. Johnson, B.-E. Pennell, I. A. L. Stoop, & B. Dorer (Eds.), *Advances in Comparative Survey Methods: Multinational, Multiregional, and Multicultural Contexts (3MC)* (pp. 325-340). Hoboken, NJ: Wileys and Sons.
Rother, N., Schacht, D., & Scheible, J. A. (2017). Sprachpotenziale: Sprachkenntnisse und Alphabetisierungsgrad von Geflüchteten. *Politikberatung kompakt, 123*, 28-40. doi:http://hdl.handle.net/10419/176772.
SOEP. (2016a). Erhebungsinstrumente der IAB-BAMF-SOEP-Befragung von Geflüchteten 2016: Haushaltsfragebogen, Stichproben M3-M4. *SOEP Survey Papers, 361*
SOEP. (2016b). Erhebungsinstrumente der IAB-BAMF-SOEP-Befragung von Geflüchteten 2016: Integrierter Personen- und Biografiefragebogen, Stichproben M3-M4. *SOEP Survey Papers, 362*
SOEP. (2018). SOEP-Core v33.1 - Documentation of Person-related Status and Generated Variables in $PGEN. *SOEP Survey Papers, 483*
Tourangeau, R., Kreuter, F., & Eckmann, S. (2015). Motivated misreporting: shaping answers to reduce survey burden. In U. Engel (Ed.), *Survey Measurements: Techniques, Data Quality and Sources of Error* (pp. 24-41). Frankfurt a. M.: Campus.
Tourangeau, R., & Smith, T. W. (1996). Asking sensitive questions: The impact of data collection mode, question format, and questions context. *Public Opinion Quartely, 60*(2), 275-304. doi:10.1086/297751

Turner, C. F., Ku, L., Rogers, S. M., Lindberg, L. D., Pleck, J. H., & Sonenstein, F. (1998). Adolescent sexual behavior, drug use and violence: Increased reporting with computer survey technology. *Science, 280*(8), 867-873. doi:10.1126/science.280.5365.867

Zavala-Rojas, D. (2018). Exploring Language Effects in Crosscultural Survey Research: Does the Language of Administration Affect Answers About Politics? *methods, data, analyses, 12*(1), 127-150. doi:10.12758/mda.2017.16

Appendix

Table A1: Commonest Mentioned Mother Tongue over Country of Origin

Country of Origin	Most common mother tongue	%
Afghanistan	Dari	62
Albania	Albanian	98
Eritrea	Tigrinya	89
Iraq	Arabic	46
Iran	Farsi	84
Kosovo	Albanian	77
Macedonia	Macedonian	46
Nigeria	English	22
Pakistan	Urdu	58
Russia	Chechen	61
Serbia	Serbian	64
Somalia	Somali	90
Syria	Arabic	73
Ukraine	Ukrainian	52
Stateless	Arabic	92

Source: IAB-BAMF-SOEP Survey of Refugees, author's calculations

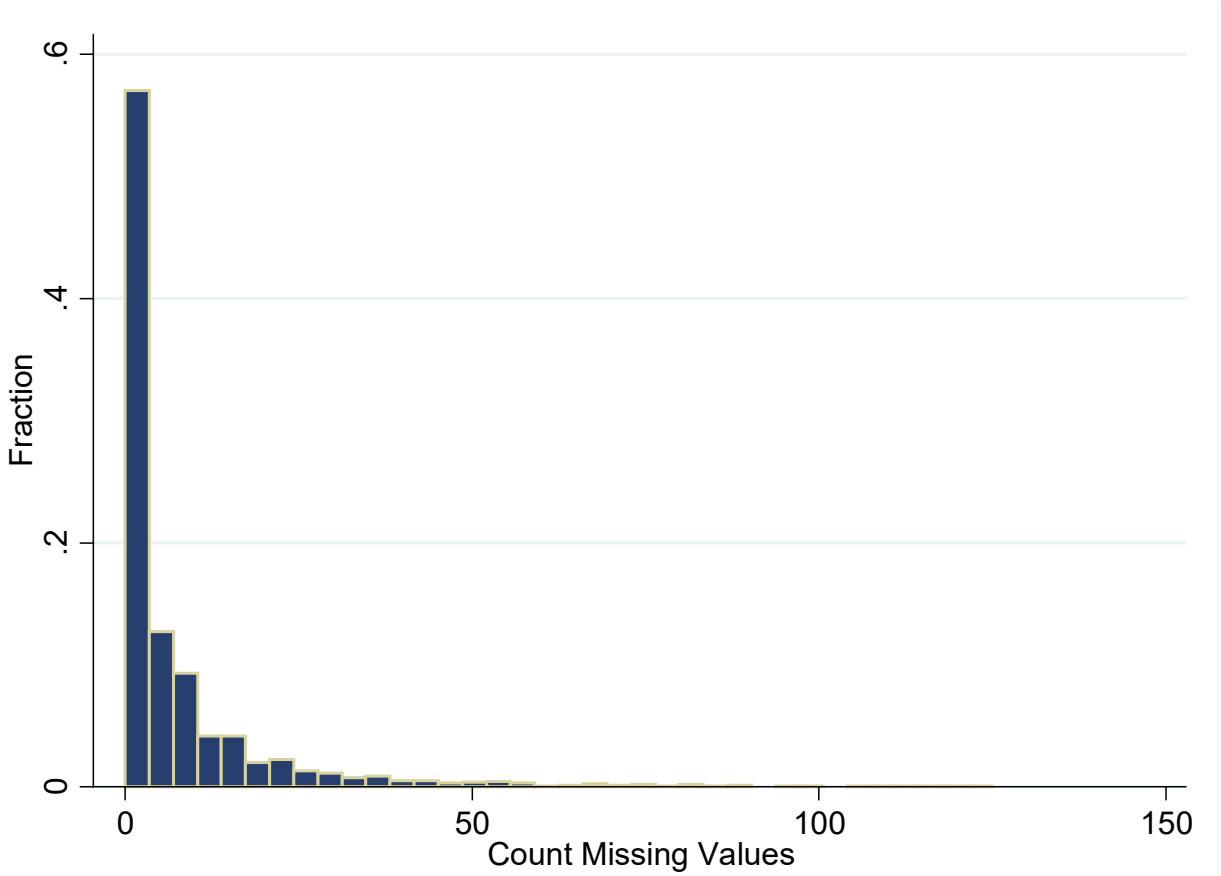

Figure A1: Histogram of Count Variable for Amount of Individually Unanswered Questions
Source: IAB-BAMF-SOEP Survey of Refugees, author's calculations

Table A2: Documentation of Tested Variables in Multilevel Logistic Regression

Model	Question	Scale	% Missing	Concept
1	Month of first immigration to Germany	All months	18.3	Immigration History
2	The direction of my life depends on me.	1 "Totally disagree" - 7 "Totally agree"	12.2	Locus of Control
3	In comparison with others, I haven't achieved what I deserved to achieve.		16.3	
4	What can be achieved in life is mainly a result of fate or luck		10.8	
5	If you are socially or politically active, you can influence social circumstances.		18.0	
6	I don't have much control of what happens in my life		10.2	
7	When I encounter difficulties, I often doubt my abilities		12.7	
8	The options that I have in life are determined by social circumstances		15.7	

90	The abilities we have are more important than the efforts we make		14.1	
10	I don't have much control over what happens in my life		12.1	
11	I think I can develop further if I deal with difficult situations		10.3	
12	I actively seek ways to balance out the losses that have affected me in my life.		10.7	
13	How well has political freedom been achieved currently in your country of origin?	0 "Very badly" - 10 "Very well"	11.3	Political Situation in Country of Origin
14	How well has civic freedom been achieved currently in your country of origin, such as the freedom to express opinions, right of assembly and an independent judiciary?		11.6	
15	How well has freedom of the press and freedom of opinion been achieved currently in your country of origin?		12.2	
16	How well is the right to practice religion or faith achieved currently in your country of origin?		11.7	
17	How well is equal treatment of ethnic minorities achieved currently in your country of origin?		14.0	
18	You need a strong leader who does not have to be concerned with a Parliament or elections	1 "Totally disagree" - 7 "Totally agree"	25.8	Forms of Government
19	Experts, not the Government, should decide what is best for the country		27.1	
20	There should be a democratic system		12.0	
21	The government taxes the rich and supports the poor	0 "Should definitely happen in a democracy" - 10 "Should definitely not happen in a democracy"	15.2	Attitudes towards democracy
22	Religious leaders ultimately determine the interpretation of laws.		20.6	
23	The people choose their government in free elections		11.9	
24	Civil rights protect the people from government oppression.		16.5	
25	Minorities are protected		15.6	

26	If a women earns more money than her partner, this inevitably leads to problems	1 "Totally disagree" - 7 "Totally agree"	10.8	Gender Equality

Source: IAB-BAMF-SOEP Survey of Refugees, author's calculations

Table A3: Documentation of Control Variables

Variable	Scale	Original Question	Remark
Same gender of interviewer and respondent?	1 = yes 0 = no	Reported by interviewer.	
ISCED	1 primary education and less 2 secondary education 3 Bachelor's and higher		See a publication by the SOEP for the generation of this variable (SOEP, 2018)
Grouped (terciles) additive index of self-perceived German proficiency (reading, writing, speaking)	1 = good 2 = medium 3 = bad	How well can you speak/write/read in German? 1 "very well" – 5 "not at all"	
Religious denomination (condensed)	1 = Islam 2 = Christianity 3 = no denomination 4 = Other	Do you belong to a church, religious community of faith? Which Christian/Islamic denomination do you belong to?	
Age in years	Metric	Your date of birth?	
Gender of respondent	1 = male 0 = female	Your Gender?	
Grouped (terciles) of duration of interview	1 = short (< 60 min) 2 = medium (< 108 min) 3 = long (< 666min)		Automatically generated by the CAPI Software
Mental distress (PHQ-4)	1 = yes 0 = no	How often have you felt negatively affected by the following complaints in the last two weeks? a) Little interest or pleasure in your activities? b) Low spirits, melancholy or hopelessness? c) Nervousness, anxiety or tension? d) Unable to stop or control worrying? 1 "not at all" – 4 "(almost) every day"	See Kroenke et al., for details and coding. (Kroenke et al., 2009)
Residence status (grouped)	1 = Asylum seeker 2 = Refugee 3 = Suspension of Deportation	Which residence title do you currently hold? If you are not sure which residence	

	4 = Other	title you hold, please check the Immigration Authority's label in your passport. a) Permission to stay pursuant to Section 55 of the German Asylum Law (asylum seekers) b) A residence permit according to Section 25 sub-section 1 of the German Residence Act (persons entitled to asylum) c) A residence permit according to Section 25 sub-section 2 of the German Residence Act (persons with refugee status) d) A settlement permit according to Section 26 sub-section 3 of the German Residence Act e) A temporary suspension of deportation according to section 60a of the German Residence Act f) A residence permit according to Section 22 or Section 23 of the German Residence Act (admission on humanitarian grounds) g) A residence permit pursuant to § 23a or § 25 sub-section 3, 4 or 5 of the German Residence Act (admission on other humanitarian grounds h) Another residence title	

Type of accommodation	1 = Shared/Public accommodation 0 = Private apartment/house	Reported by the interviewer as answer to: In what type of accommodation does the interviewee live?	
Grouped (terciles) additive index of perception of democratic freedom in home country	1 = low 2 = medium 3 = high	How would you personally assess the current situation in your country of origin regarding the following areas of live. How well has 1) Political Freedom 2) Free speech 3) Freedom of press 4) Religious freedom 5) Equality of minorities 6) Equality of men and women been achieved currently in your country of origin. Scale: 0 “Very bad” to 10 “very good”	
Bystander during interview	1 = yes 0 = no	Reported by the interviewer	
Functional illiteracy	1 = yes 0 = no	How well can you write/read in your native language? 1 “very well” – 5 “not at all”	Respondents with low (4 or 5 on the scale) abilities to read and write in their mother tongue.

Also see the publicly available questionnaires of the SOEP group (SOEP, 2016a, 2016b). A translation from German to English is included.

Table A4: Frequencies of Control Variables used in multivariate models

Variable	Share in %
Same gender of interviewer and respondent?	
yes	40.1
no	59.9
ISCED	
Primary and lower	38.6
Secondary	41.9
Tertiary	19.5
Grouped additive index of self-perceived German proficiency (reading, writing, speaking)	
Good	20.6
Intermediate	33.1
Bad	46.4
Religious denomination (condensed)	
Islam	70.7
Christian	14.6
No denomination	6.9
Other	7.8
Age in years	Mean = 33.6 / SD = 10.4
Gender of respondent	
Female	36.7
Male	63.3
Grouped (terciles) of duration of interview	
Short	21.8
Medium	49.5
Long	28.8
Mental distress (PHQ-4)	
No	67.1
Yes	32.9
Residence status (grouped)	
Asylum Seeker	32.2
Residence Permit	55.7
Suspension of Deportation	7.0
Other	5.2
Type of accommodation	
Private	68.3
Shared	31.7
Grouped additive index of perception of democratic freedom in home country	
Low (lowest quartile)	24.5
Medium	50.6
High (Highest Quartile)	24.9
Bystander during interview	
No	42.5

	Yes	57.5
Functional illiteracy		
	No	87.1
	Yes	12.9
N = 3,630		

Listwise deletion of missing values

Table A5: Zero Inflated Poisson Regression Analysis with Control Variables

	Zero Inflated Poisson Regression (ZIP) on number of unanswered Questions [a b] (unstandardized coefficients) SE in parenthesis					
Independent Variables	Model 1a			Model 1b		
Match of mother tongue, ref = no	-.30	***	(.02)	-.31	***	(.02)
Audio recording used, ref = not used	.14	***	(.03)	.08		(.05)
Interaction Audio recording*Match	-	-	-	.08		(.08)
Same Gender between Interviewer and Interviewee, ref = no	-.01		(.02)	-.01		(.01)
ISCED, ref = primary and less						
Secondary	-.20	***	(.02)	-.43	***	(.09)
Tertiary	-.34	***	(.03)	-.57	***	(.09)
German Proficiency, ref = good						
Intermediate	.05	**	(.03)	.05	**	(.03)
Poor	.07	***	(.03)	.08	***	(.03)
Religious Denomination, ref = Islam						
Christ	.09	***	(.02)	.09	***	(.02)
No Denomination	.01		(.03)	.01		(.03)
Other	.07	**	(.03)	.07	**	(.03)
Age	-.01	***	(.00)	-.01	***	(.00)
Gender, ref = female	-.12	***	(.02)	-.12	***	(.02)
Duration of Interview, ref = short						
medium	.10	***	(.02)	.10	***	(.02)
long	.21	***	(.02)	.21	***	(.02)
Mental distress, ref = no	.08	***	(.02)	.09	***	(.02)
Residence Status, ref = Asylum Seeker						
Residence Permit	-.03		(.02)	-.03		(.02)
Suspension of Deportation	-.09	***	(.03)	-.09	***	(.03)
Other	.03		(.04)	.02		(.04)
Type of Accommodation, ref = Private	.02		(.02)	.02		(.02)
Perceived Level of Democracy in Home Country, ref = low						
Medium	-.10	***	(.02)	-.10	***	(.02)
High	.12	***	(.02)	.12	***	(.02)
Bystander, ref = no	.04	**	(.02)	.04	**	(.02)
Functional Illiteracy, ref = no	.15	***	(.02)	.15	***	(.02)
Intercept	2.20	***	(.10)	2.43	***	(.10)
Chi^2	1472			1479		

Source: IAB-BAMF-SOEP Survey of Refugees, author's calculations
***p<0.01 **p<0.05 *p<0.1 (two-tailed-tests)
N (all models) = 3,630
Nonzero Observations = 2,505

Table A6: Robustness Check: Interaction Term of Language Match and Interview Language in Poisson Regression Analysis with Count of Item-Nonresponse per Individual as Dependent Variable.

	Zero Inflated Poisson Regression (ZIP) on number of unanswered Questions [a] (unstandardized coefficients) SE in parenthesis		
Independent Variables			
Match of mother tongue, ref = no	-.75	***	(.22)
Interview Language, ref = English			
Arabic	-.42	***	(.03)
Farsi/Dari	-.65	***	(.08)
Pashto	.01		(.18)
Urdu	-.27	***	(.09)
Kurmanji	-.25	*	(.13)
Interaction Match/Language, ref = English			
Arabic	.52	**	(.22)
Farsi/Dari	.84	***	(.24)
Pashto	.62	**	(.29)
Urdu	.70	***	(.26)
Kurmanji	.74	***	(.26)
Audio recording used, ref = not used	.13	***	(.03)
Same Gender between Interviewer and Interviewee	-.01		(.02)
ISCED, ref = Primary and lower			
Secondary	-.20	***	(.02)
Tertiary	-.32	***	(.03)
German Proficiency, ref = good			
Intermediate	.05	*	(.03)
Poor	.08	***	(.03)
Religious Denomination, ref = Islam			
Christ	-.04	*	(.02)
No Denomination	.04		(.04)
Other	.04		(.04)
Age	-.01	***	(.00)
Gender, ref = female	-.12	***	(.02)
Duration of Interview, ref = short			
medium	.10	***	(.02)
long	.20	***	(.02)
Mental distress, ref = no	.07	***	(.02)
Residence Status, ref = Asylum Seeker			
Residence Permit	.05	***	(.02)
Suspension of Deportation	-.19	***	(.03)
Other	.07	*	(.04)
Type of Accommodation, ref = Private	.00		(.02)
Perceived Level of Democracy in Home Country, ref = low			
Medium	-.05	**	(.02)

High	.19	***	(.02)
Bystander, ref = no	.05	***	(.02)
Functional Illiteracy, ref = no	.14	***	(.02)
Intercept	2.37	***	(.10)
Chi^2	1868		

Source: IAB-BAMF-SOEP Survey of Refugees, author's calculations
***p<0.01 **p<0.05 *p<0.1 (two-tailed-tests)
N = 3,630
Nonzero Observations = 2,505
[a] The estimates of the Poisson part of the function are presented only. The estimation of having no nonresponse is available upon request.

Table A7: Linear Probability Regression Analysis with Control Variables

	Linear Probability Regression Analyses with Random Intercepts [a] (Average Marginal Effect, 1 = Nonresponse) SE in parenthesis					
Independent Variables	Model 1a			Model 1b		
Match of mother tongue, ref = no	-.05	***	(.01)	-.05	***	(.01)
Audio recording used, ref = not used	.05	***	(.01)	.01		(.02)
Interaction Audio recording*Match	-	-	-	.06	**	(.03)
Topic, ref = Month of Immigration						
Locus of Control	-.06	***	(.00)	-.06	***	(.00)
Political Situation in CoO	-.07	***	(.00)	-.07	***	(.00)
Forms of Government	.03	***	(.00)	.03	***	(.00)
Attitudes towards Democracy	-.03	***	(.00)	-.03	***	(.00)
Attitudes towards Gender Equality	-.08	***	(.01)	-.08	***	(.01)
Same Gender between Interviewer and Interviewee	.01	*	(.01)	.01	*	(.01)
ISCED, ref = Primary and lower						
Secondary	-.04	***	(.01)	-.04	***	(.01)
Tertiary	-.07	***	(.01)	-.07	***	(.01)
German Proficiency, ref = good						
Intermediate	.00		(.01)	.00		(.01)
Poor	.02	***	(.01)	.02	***	(.01)
Religious Denomination, ref = Islam						
Christ	.04	***	(.01)	.03	***	(.01)
No Denomination	-.01		(.01)	-.01		(.01)
Other	-.01		(.01)	-.00		(.01)
Age	-.00	***	(.00)	-.00	***	(.00)
Gender, ref = female	-.03	***	(.01)	-.03	***	(.01)
Duration of Interview, ref = short						
medium	-.00		(.01)	-.00		(.01)
long	.00		(.01)	.00		(.01)
Mental distress, ref = no	.00		(.01)	.00		(.01)
Residence Status, ref = Asylum Seeker						
Residence Permit	-.02	**	(.01)	-.02	***	(.01)
Suspension of Deportation	.02		(.01)	.02		(.01)
Other	.02		(.01)	.02		(.01)
Type of Accommodation, ref = Private	-.00		(.01)	-.00		(.01)
Bystander, ref = no	-.01	**	(.01)	-.01	**	(.01)
Functional Illiteracy, ref = no	.06	***	(.01)	.06	***	(.01)
Intercept	.26	***	(.04)	.26	***	(.04)

Source: IAB-BAMF-SOEP Survey of Refugees, author's calculations
***p<0.01 **p<0.05 *p<0.1 (two-tailed-tests)
N = 97,266, N(Respondents) = 3,741, N(Interviewer) = 79

Table A8: Mean Probability for Questions With More Than 10 Percent Missing Values to Remain Unanswered by Respondents with Different Mother Tongues

Mother Tongue	Mean response probability in percent [95 percent Confidence Interval]
Arabic[1]	8 [7.7-7.8]
Urdu[1]	10 [9.3-10.1]
Farsi[1]	12 [11.4-11.8]
Dari[1]	12 [12.4-12.5]
Pashto[1]	12 [11.9-12.7]
Kurdish-Kurmanji[1]	14 [14.3-14.5]
Russian	14 [13.9-14.6]
Punjabi	14 [14.1-14.9]
Albanian	15 [15.0-15.4]
Armenian	15 [14.8-15.5]
Kurdish-Sorani	15 [14.8-15.6]
Somali	15 [15.1-15.8]
Aramaic	16 [15.5-16.4]
Kurdish-South Kurdish	17 [15.9-17.6]
Tigrinya	18 [17.5-17.8]
Serbian	18 [17.4-18.1]

[1]Provided visual translation

Source: IAB-BAMF-SOEP Survey of Refugees, author's calculations for countries with N >= 20 in IAB-BAMF-SOEP Survey of Refugees based on model 1a in Table 5

Note: As linear probability models can produce estimated probabilities outside 0,1, all estimates above and under are trimmed to 0 and 1 respectively (for 3% of all estimates).

Table A9: Robustness Check: Detailed Usage of Audio Recordings and Its Effects on the Amount of Unanswered Questions

	Zero Inflated Poisson Regression (ZIP) on number of unanswered Questions [a] (unstandardized coefficients) SE in parenthesis		
Independent Variables			
Match of mother tongue, ref = no	-.30	***	(.02)
Audio Recordings, ref = used with every question			
with 2/3 of the questions	-.11	***	(.04)
with around half of the questions	-.06		(.05)
with less than half of the questions	-.20	***	(.04)
not at all	-.14	***	(.03)
Same Gender between Interviewer and Interviewee	-0.02		(.01)
ISCED, ref = Primary and lower			
Secondary	-.20	***	(.02)
Tertiary	-.34	***	(.03)
German Proficiency, ref = good			
Intermediate	.05	**	(.03)
Poor	.07	***	(.03)
Religious Denomination, ref = Islam			
Christ	.09	***	(.02)
No Denomination	.01		(.03)
Other	.07		(.03)
Age	-.01	***	(.00)
Gender, ref = female	-.12	***	(.02)
Duration of Interview, ref = short			
medium	0.1	***	(.02)
long	.21	***	(.02)
Mental distress, ref = no	.08	***	(.02)
Residence Status, ref = Asylum Seeker			
Residence Permit	-.03		(.02)
Suspension of Deportation	-.09	**	(.03)
Other	.02		(.04)
Type of Accommodation, ref = Private	.01		(.02)
Perceived Level of Democracy in Home Country, ref = low			
medium	-.10	***	(.02)
high	0.12	***	(.02)
Bystander, ref = no	0.04	**	(.02)
Functional Illiteracy, ref = no	.14	***	(.02)
Intercept	2.34	***	(.06)
Chi^2	1482		

Source: IAB-BAMF-SOEP Survey of Refugees, author's calculations
***p<0.01 **p<0.05 *p<0.1 (two-tailed-tests)
N = 3,630
Nonzero Observations = 2,505

[a]The estimates of the Poisson part of the function are presented only. The estimation of having no nonresponse is available upon request.

6 Article 3: Conceptions of Democracy in Cross-Linguistic and Cross-National Research

Preface

The first two papers deal with different forms of selection bias arising either by design (sampling) or due to self-selection (item nonresponse). As discussed, both aspects pose problems to generalizability. However, if the selection criteria are known or can be estimated, this can be corrected for. But, even if this is done, there are further problems to generalizability. One of these is measurement invariance. Measurement invariance is the prerequisite for between-group comparisons across time or between different societal groups. The concept assesses whether those groups of interest have the same understanding of a latent concept measured in questionnaires. As refugees are an extremely diverse target population (in terms of language, socialization, and cultural history), and as many scholars are interested in comparisons between refugees and local populations, measurement invariance is a crucial aspect for a survey of recent refugees.

When applying the concept of measurement invariance to integration research, it becomes clear that latent constructs as markers of integration necessarily have to be tested across groups. Prominent in this regard are attitudes towards democracy, psychosocial scales (e.g., well-being, PTSD), or the measurement of language proficiency. As especially the latter two aspects have already been the subject of much debate, the following paper takes the understanding of democracy as an example in order to show what factors can impede comparability.

Can We Compare Conceptions of Democracy in Cross-Linguistic and Cross-National Research? Evidence from a Random Sample of Refugees in Germany

The article is published and should be cited as follows:
Jacobsen, Jannes; Fuchs, Lukas M. (2020) Can We Compare Conceptions of Democracy in Cross-Linguistic and Cross-National Research? Evidence from a Random Sample of Refugees in Germany. *Social Indicators Research.* DOI: 10.1007/s11205-020-02397-6

Abstract

This study addresses the heated academic and public debate on the compatibility and comparability of refugees' and host societies' democratic values. Comparative values research has long capitalized on global similarities and differences in support for Western democratic values. We argue that such cross-cultural comparisons of culturally diverse groups are challenged by 1) different conceptions of democracy determined by different experiences with democratic systems and 2) bias introduced by linguistic differences and translation processes. In order to analyze whether the conception of democracy is comparable between different nationalities and languages, we test data from the German IAB-BAMF-SOEP Survey of Refugees and the World Values Survey (WVS) for measurement invariance using multi-group confirmatory factor analysis (MGCFA). Our results suggest that the applied democracy scales are problematic for comparing conceptions of democracy between refugees and Germans. Furthermore, we show that such conceptions are also not comparable across mother tongues and survey language.

1. Introduction

With the unprecedented influx of refugees to Europe between 2015 and 2018, an intense public debate arose in Germany over how to accommodate and integrate the new arrivals. One crucial aspect of this debate was the fear that the newcomers do not share fundamental values of the host society (Banulescu-Bogdan & Benton, 2017). Studies began to address the question of how to convey German values to the newly arriving refugees and how to measure their agreement with these values (Banulescu-Bogdan & Benton, 2017; Müller-Hilmer & Gagné, 2018; SVR, 2019). From a political and sociological perspective, this question is important for two reasons. First, value consensus—a group's collective agreement with certain fundamental ethics and ideals (Parsons, 1968; Wan, Chiu, Tam, Lee & Lau, 2007)—is said to enhance social cohesion by promoting cooperation and simultaneously preventing conflict (Partridge, 1971). Following this line of argumentation, it should be possible to predict conflict or cooperation between immigrants and the receiving society by assessing whether the two groups share the same values. Second, research on comparative values deals with questions of whether values measured by means of quantitative methods are actually comparable between different cultures, that is, both between and within countries. If values are not comparable between cultures, it would be almost impossible to analyze whether people share the same values and thus also whether immigrants threaten certain values of the host society.

In the public and political debate in Germany and Europe, liberal democratic values are often conflated with European or national values and have been studied extensively. In this research, they are described as the foundation of stable liberal democracies because they mirror the aspiration to support and be actively involved in political processes, the central arena of societal participation (Diamond & Linz, 1989; Shin, 2007). The use of inferential statistics in analyzing and comparing democratic values has long been a focus of interest in comparative values studies producing a vast body of literature comparing democratic values between countries or cultural entities: the results served as proxies for the democratic condition of a state and the chances that a country will become (or remain) democratic (Diamond & Plattner, 2015; Linz & Stepan, 1996). Outside academia, results from such general population surveys tend to be used to fuel those concerns pertaining to refugees and immigrants as threats to European or "Western" values. Such assumptions are rash for several reasons: First, comparisons from general population surveys, cannot be extrapolated to highly selective samples of migrants and refugees and are thus unfit to draw conclusions on whether or not refugees challenge European democratic values. Second and more importantly, more recent research strongly warrants a more cautious approach to comparing complex concepts like values and attitudes, indicating that especially attitudes toward democracy in particular are at

risk of not being comparable across political cultures (e.g., Ariely & Davidov, 2011). Numerous recent studies criticize comparative values research for often ignoring or failing to establish necessary levels of measurement equivalence (Davidov, Meuleman, Cieciuch, Schmidt & Billiet, 2014), a precondition for any study involving cross-cultural comparison of democratic values (Canache, 2012).

Focusing on the influx of refugees into Germany as a case study this article expands the current academic debate on measurement equivalence, demonstrating that even in a study that is conducted in the same cultural and historical context (Germany) careful assessment of whether conceptions of democracy are comparable amongst refugee respondents from diverse backgrounds is mandatory. In a number of publications on this topic, such tests are lacking (e.g., Brücker et al., 2016 use a German panel study on refugees and pool it with the WVS; Buber-Ennser et al., 2016), rendering the empirical and statistical comparisons of conceptions of democracy between respondents of different origins and their conclusions about their democratic values flawed. We develop the theoretical argument that even in surveys that are conducted in a single national context, two main factors may pose a challenge to comparability. The first of these is experience with democratic systems. Different countries of origin are here considered indicators for people's past experience with democratic systems. Because there is no clear benchmark definition of democracy, different conceptions of democracy may exist in different populations. The definition of what constitutes a democracy is therefore likely to vary among asylum seekers from diverse and often quite undemocratic backgrounds (a similar argument is made by Ariely & Davidov, 2011). Second, from the standpoint of survey methodology, the respondent's language poses additional challenges to comparability (Zavala-Rojas & Saris, 2017). This is due both to the fact that words and semantics often do not translate directly from one language to another (Bratton & Mattes, 2001), and to the bias introduced by translation itself (Behr, Brzoska & Schoua-Glusberg, 2018; Comanaru & d'Ardenne, 2018; Goerman, Meyers & Trejo, 2018).

We explore both of the aforementioned challenges by analyzing data from the German IAB-BAMF-SOEP Survey of Refugees a unique sample of refugees in Germany and later pooling it with the German World Value Survey (WVS). We start with an overview of previous comparative research on democratic conceptions and values, illustrating that much of the interpretation to date has been biased by insufficient assessment of measurement invariance. In a second step, we revisit the ongoing academic debate on measurement equivalence and bring forward two theoretical arguments for why we consider measurement invariance imperative when studying a sample of nationally and culturally diverse refugees. In the

empirical section, by testing measurement invariance (for an example of the method see Saris, Pirralha & Zavala-Rojas, 2018), we indicate that democracy remains an ambivalent concept amongst individuals from different political cultures and with different mother tongues and that cross-cultural and cross-linguistic comparisons are likely to be problematic. Hence, in the sensitive context of refugee and migration research a comparison of democratic values between different political cultures using quantitative methods has to be supported by careful considerations of measurement invariance.

2. Lessons Learned From Comparative Values Research

Cultural values are among the most prominent areas of sociological research, not least for measuring possible cultural diversification in the wake of transnational migration. Values are known to assume the role of mediators between individual conceptions of the desirable and undesirable, on the one hand (Marini, 2000; see also Kluckhohn, 1951), and societal demands, on the other hand (Grube, Mayton & Ball-Rokeach, 1994). They thus govern societal cooperation by defining ideal modes of interaction and coexistence as well as determining the conditions for conflict settlement. Agreement to values that foster cooperation and prevent conflict in particular is described as the foundation of liberal civil democracies (Inglehart & Welzel, 2005). Democratic values here are of specific prominence as expressions of an individual's satisfaction with society and at the same time, as determinants of the stability of democratic institutions and systems (Inglehart, 2000). Empirical research on cross-cultural agreement to democratic and societal values has produced diverse and in part conflicting results to date (for an overview see also Gabriel, 2020).

In times of high transnational migration, studies on global differences in values, particularly those providing comparisons between "Western" and "Non-Western" societies, are prominent: Many studies map the globally changing support for the democratic values items in the World Value Survey (WVS) from a historical perspective (e.g. Welzel, 2013). The main argument is that democratic enhancements can be causally explained by socio-economic development and increasing prosperity (Inglehart & Baker, 2000; Inglehart & Welzel, 2005; Inglehart & Welzel, 2009). While this narrative seemingly works on an overall scale, it accounts primarily for developments in Western countries. Looking closer, its explanatory power seems limited in explaining the strongly diverging effect sizes of economic development on democratic or equality values between "Western" and "Non-Western" societies.

A competing body of studies, though similar in authorship, thus capitalizes on the cross-cultural differences in democratization, secularism, and gender equality (Inglehart & Welzel, 2009; Norris & Inglehart, 2012). These studies argue that there is in fact no globally increasing

support for democracy and instead accentuate differences in democratic support between what they categorize as liberal, secular Western countries, and clerical, patriarchal non-Western societies, i.e. between refugee sending and receiving societies (Alexander & Welzel, 2011; Tausch & Heshmati, 2003; Welzel, 2013). Inglehart and Norris (2003), for example, concluded that "Muslims and their Western counterparts" desire democracy equally, but at the same time, that Muslims do not share Western egalitarian and equality values. Instead of investigating how values surveys find simultaneously that Muslim and Western societies differ little in their desire for democracy but differ strongly in their conceptions thereof, they translated these findings into a generally pessimist outlook for democracy in Muslim countries. In the same vein, Alexander and Welzel (2011) argued that Muslim support for patriarchal values is robust across time as well as geographic space, irrespective of democratic advancements, vaguely blaming religious and cultural factors but not empirical approaches to democracy research.

These already inconclusive findings, however, have often been more or less directly conferred onto migrants from Non-Western countries, refugees in particular, to foresee cultural clashes and value conflicts (Tausch, 2016). Such conclusions are hasty for numerous reasons: First, migrants and refugees are usually a highly selective group compared to those who stay behind and should not be considered representative for their countries of origin (Belot & Hatton, 2012; Docquier, Tansel & Turati, 2018; Wimmer & Soehl, 2014). Second, comparative values studies on a global scale tend to overestimate value homogeneity within countries (Schwartz & Sagie, 2000). Within-country variations in democratic values are in fact often stronger than the aggregate differences between countries (Silver & Dowley, 2000; Fischer & Schwartz, 2011). Finally and most far-reaching, democratic values and conceptions of democracy in refugee countries of origin are likely to differ from those in Western Europe. Using the Arab Barometer, Kostenko et al. (2016) for example demonstrated that democratic values in Arab countries are not linked to gender equality (Kostenko, Kuzmuchev & Ponarin, 2016; Rizzo, Abdel-Latif & Meyer, 2007). Meanwhile, Vlas and Gherghina (Vlas & Gherghina, 2012), contest claims about Muslim patriarchy by showing that democratic and equality values are not linked to religion but rather to living in a patriarchal society.

Comparative values studies—despite longstanding and extensive research—in some regards produced controversial and in parts inconclusive results. We argue that comparative values studies have often suffered from an empirical bias and lack of rigor assessments of comparability of value conceptions. This can be particularly harmful where this engenders hasty conclusions concerning potentially salient areas, such as refugee accommodation. In

recent years, however, a growing body of research addressing this issue emerged. The following section discusses these recent developments.

2.1 Testing Perceptions and Attitudes Towards Democracy for Comparability: Recent Findings

From the perspective of empirical and survey-based social science research, before comparing conceptions of democracy or democratic values it is crucial to ask whether the underlying concept of democracy is comparable, meaning that people actually think about the same concept when hearing the term democracy. Only if this is the case comparisons are unbiased. Yet, two seemingly opposing camps are involved in an ongoing academic debate as to how measurement invariance should be assessed – strict proponents of testing constructs' internal measurement invariance on the one hand (e.g. Ariely & Davidov, 2011) and those championing constructs' external and aggregate validity (e.g. Welzel & Inglehart, 2016).

The majority of studies assessing the measurement invariance of democratic values comes to the conclusion that there are major differences, both cross-culturally and cross-nationally. Using the Arab Barometer, Tessler, Jamal, and Robbins (2012) used a novel approach to estimate differences in perceptions of democracy. They asked respondents from Algeria, Jordan, Lebanon, Palestine, Yemen, Egypt, and Tunisia for their understanding of democracy. The response options were "free elections", "freedom of speech", "low economic inequality", and "basic necessities for all". Their results reveal that none of the potential outcomes are mentioned by more than thirty percent of respondents, which gives a strong indication that these populations hold diverse understandings of democracy. Likewise, Ariely and Davidov (2011), using WVS data and confirmatory factor analysis, question that concepts such as "democracy-autocracy preference" (DAP) and "democratic-performance evaluation" (DPE) are comparable cross-nationally. For the DAP they find that although the understanding of the items might be similar, comparing means is problematic. At the same time, they find that the DPE means are comparable across a large set of countries. Meanwhile, Behr et al. (2014) assessed ISSP Data to demonstrate that the "civil disobedience" item as part of the "rights in a democracy" is understood different in the United States and Canada in contrast to European countries Denmark, Germany, Hungary, and Spain. Using the Latin Barometer and survey data from Romania, Canache, Mondak, and Seligson (2001) showed that the measurement of the well-known satisfaction with democracy (SWD) concept is not reliable cross-nationally (Canache, Mondak and Seligson 2001; Linde & Ekman, 2003).

Opposed to this more strict and technical approach, a counter movement led by Welzel and Inglehart (2016) argues that measurement invariance tests have fetishized a construct's

internal validity without regards to potential external validation. Welzel and Inglehart (2016) indicate that amid careful theoretical considerations democratic values might nevertheless be comparable cross-culturally. And indeed, in a 2012 study, Ariely and Davidov reported that the "attitudes towards government intervention" scale of the ISSP is comparable between the United States, Britain, West Germany, and Sweden. Other studies tried to find methodological solutions to the challenge of measurement equivalence in democracy research. Schedler and Sarsfield (Schedler & Sarsfield, 2007) proposed the use of cluster analysis to study different conceptions of democracy. Using the Mexican 2003 National Survey on Political Culture, they showed that although there is general support for democracy, people can be divided in different groups reflecting deeper understandings of democracy. In a recent study, Ulbricht (2018), while showing that the understanding of democracy indeed varies around the world and that support for representative democracy has been substantially overestimated in previous research, maps out an innovative Analytical Hierarchy Process that allowed him to assess different conceptions of democracy (Ulbricht, 2018).

In light of the inconclusive ongoing debate and the contradictory findings concerning democratic values' comparability, we argue that for research on delicate topics, such as refugee's value conceptions, testing for measurement invariance needs to be a precondition. Otherwise, the debate on the contestation of Germany's social cohesion and democratic condition as a receiving country is prone to be misguided by faulty data. Based on previous research on value conceptions, we identify two aspects that are likely to hamper comparability: political culture and language.

2.2 Democracy: A Cross-Culturally Ambiguous Concept

Bueno (2012) as well as Ariely and Davidov (2011) argue that the absence of comparability between different countries is the result of different political cultures (Ariely & Davidov, 2011; Bueno, 2012). If people have different experiences with democracy from one country to another, their perceptions of democracy must be diverse as well. Understanding that conceptions of democracy are strongly shaped by the cultural and historical context points to the "paradox of democracy" (Alvarez & Welzel, 2014), the idea that support for democracy in a given country does not reflect the actual democratic state of that country. Support for democracy is argued to be linked primarily to the cognitive understanding of democracy and knowledge of institutional functioning (Miller, Hesli & Reisinger, 1997). The relation between people's awareness of and support for democracy is, however, not linear but instead mediated and influenced by individual biographical experiences with democracy (Cho, 2014). These experiences, and in turn also the understanding of democracy, are determined first and

foremost by the cultural context and educational system in the country of origin (Finkel & Smith, 2011). This gives rise to our first hypothesis:

Hypothesis 1: Conceptions of democracy are not comparable between refugees from different countries and the local German population.

2.3 Linguistic Challenges in Cross-Cultural Democracy Research

Linguistic and cross-cultural research shows that concepts which are referred to by the same name can still vary between languages, cultures, and states (Behr et al., 2018). Thus, in addition to the aforementioned difficulties in comparing perceptions of democracy between countries, there is a second dimension challenging comparability. As most of the articles cited above use multi-lingual survey data, the aspect of questionnaire language becomes a crucial one. Translating questionnaires entails a serious risk of bias: conveying a specific meaning from one language to another is not always straightforward (Smith, 2003) and can trigger a change in attitudes (Zavala-Rojas, 2018). Some languages, for instance, have various words for a given concept, whereas others have only one. Words for democracy have entered some languages (e.g., in Africa) only very recently (Bratton & ,Mattes 2001). Furthermore, a given language can have different dialects, and people who speak the same language often use different expressions in their various dialects. A prominent example is Arabic. Although standard Arabic exists as a language, most people speak regional dialects. Thus, the formal or official language does not necessarily represent a respondent's mother tongue (Comanaru & d'Ardenne, 2018). If this causes respondents to understand questions differently, the measurement would no longer be comparable. Some concepts or terms also have different meanings in a given language, or in some cases translations do not exist and a term can only be described instead of being translated. The language itself incorporates the meaning of a term. Thus, this meaning can vary between languages and impede comparability (Davidov & Beuckelaer, 2010). This leads to our second hypothesis:

Hypothesis 2: Perceptions of democracy are not comparable across languages.

In sum, we built the theoretical argument that different political culture and language hamper the comparability of perceptions of democracy across culturally distinct samples. We therefore assume that respondents' experience with democracy and respondents' language pose a challenge to measurement invariance. These challenges are especially important when analyzing and comparing refugees, who, rather than constituting a homogenous group, are characterized by immense cultural and linguistic diversity and a variety of backgrounds (Dustmann, Fasani, Frattini, Minale & Schönberg, 2017).

3. Methods

In order to test whether the conceptions of democracy are comparable between different nationalities and languages, we test for measurement invariance and conduct multi-group confirmatory factor analysis (MGCFA) with a bottom-up stepwise procedure. This is a commonly accepted method (Medina, Smith, & Long, 2009; Saris et al., 2018; Vandenberg & Lance, 2000). Measurement invariance assures that mean differences in latent variables between groups are not due to different factor loadings or intercepts and thus meaningful comparisons can be carried out.

Generally speaking, the relationship between democratic values (ξ) and the manifest variables of conceptions of democracy as responses (y) can be described as a function of

$$y = \tau + \lambda\xi + \delta \text{ (1)}.$$

In this case the intercepts (τ) and slopes (λ) are assumed to be equal across people with e.g. different nationalities. In order to test whether the concept of democracy (DEM) is actually comparable, the equation needs to be estimated separately for each manifest variable that measures democratic values (G = {1,2,…,k}) by means of:

$$y_1 = \tau_1 + \lambda_{11} DEM + \delta_1 \text{ (2)}$$

$$y_2 = \tau_2 + \lambda_{21} DEM + \delta_2 \text{ (3)}$$

$$y_k = \tau_k + \lambda_{k1} DEM + \delta_k \text{ (4)}.$$

Further, we assume that

$$Covariance(DEM, \delta_i) = 0, for\ all\ i \text{ (5)}$$

$$Covariance(\delta_i, \delta_j) = 0, for\ all\ i \neq j \text{ (6)}.$$

We handle missing data by employing full information maximum likelihood estimation (Schafer & Graham, 2002). The models for the different manifest variables (1,2,…,k) are tested with the same restrictions simultaneously for all groups using the lavaan package implemented in R (Jöreskog, 1971; Muthen & Satorra, 1995; Rosseel, 2019).

In a first step, we assess whether the latent construct exists in all sub-groups separately but with similar configuration (configural invariance). In order to do so, the factor loadings need

to be adequate in all groups[1]. Additionally, the fit indices should not indicate a bad model fit (CFA ≥ .95; RMSEA ≤ .05). In the next two steps, we restrict the confirmatory model increasingly and test for metric and scalar invariance. At first, we restrict the factor loadings to be equal across groups (metric invariance), and second, we also restrict the intercepts to be equal across groups (scalar invariance). Between those two steps, the fit indices need to be assessed. The restrictions are commonly confirmed as adequate using the Comparative Fit Index (CFI). However, if the comparative fit index (CFI) is substantially lower than .95 or drops by more than .01, the procedure needs to be stopped (Chen, 2007; Cheung & Rensvold, 2002)[2]. In this case, the literature proposes testing the initial step again, but instead of restricting all parameters for all variables, estimating parameters for one factor freely (the variable should be determined by considering modification indices, not displayed as tables). If the assumption then holds, we might speak of partial measurement invariance. How many parameters can be estimated freely is the subject of an intense debate in the literature dealing with measurement invariance. Steenkamp and Baumgartner (1998) summarize the debate and argue that due to the nature of a confirmatory model, estimating parameters freely should be treated with caution in order to avoid applying excessive researcher's degrees of freedom. However, they also indicate that under some circumstances, restricting parameters for two variables only can be sufficient. If models rely on only few groups with quite different sample sizes and an overall only medium total sample – like in our case (compared to other studies applying CFA, e.g., Ariely & Davidov, 2011; Alemán & Woods, 2016) – literature warrants a more cautious approach (McNeish, An & Hancock, 2017). Moreover, Chen (2007) indicates that smaller samples have a higher chance of producing acceptable confirmatory models. This should be kept in mind when examining the fit indices. We therefore choose a conservative strategy and argue that at least half of the parameters should be fixed in order to make sure that the latent constructs are robust to differences in slopes and intercepts between groups, while also discussing how more liberal cut-off criteria would influence the results in our limitations section.

[1] There is not a commonly defined cut-off criterion. We assume that a factor loading is inadequate when, compared to other items, its variance is explained to a lesser degree by the latent variable. We additionally rely on the fit indices in the event that some factor loadings appeared to be substantially smaller than others.

[2] In regard to the stepwise procedure there is no clear cut-off criterion defined in the literature. Most studies however use a CFI between 0.90-0.95 (Hu & Bentler, 1999; Marsh, Tau & Wen, 2004). Therefore, in order to determine invariance, the deterioration and the absolute CFI have to be taken into account, equally. Moreover, simulation studies suggest, that models based on medium sized factor loadings should be treated more strictly (McNeish, An & Hancock, 2017). As we will present further down, many of our factor loadings are around .5, .6 and some even around .4, suggesting the application of strict and conservative thresholds.

3.1 Data

In order to have a dataset consisting of a sufficient number of Germans and recent immigrants, we pool two datasets that both employed a set of the same variables regarding conceptions of democracy: the 2016 wave of the IAB-BAMF-SOEP Survey of Refugees and the 2014 wave of the World Value Survey. The IAB-BAMF-SOEP Survey of Refugees is a random sample of refugees and asylum seekers who arrived in Germany between 2013 and 2016 (Kühne, Jacobsen, & Kroh, 2019). The World Value Survey (WVS) is a global survey on public opinions and covers around 80 percent of the world population. Separate random samples are drawn for each participating country (Inglehart et al., 2014). Both surveys employ the same four questions asking about conceptions of democracy (see Table 1). The only difference is that the WVS relies on a ten-point scale (from 1 "should definitely not happen in a democracy" to 10 "should definitely happen in a democracy") and the SOEP on an 11-point scale (0-10). In order to harmonize the scales, we split the middle category of the IAB-BAMF-SOEP scale randomly between the neighboring steps.

In order to test whether conceptions of democracy are comparable between refugees and the German population, we use the four largest national groups in the IAB-BAMF-SOEP Survey of Refugees. Excluding all other countries from the refugee sample is necessary because they are not represented by a sufficient number of respondents. German respondents are identified in the WVS and integrated to the refugee survey (see Table 2).

Table 1: Documentation of Manifest Variables

Manifest Variable	**Dataset**
Do you think that the following things are what should happen in a democracy or not?	
The government taxes the rich and supports the poor	IAB-BAMF-SOEP/WVS
Religious leaders ultimately determine the interpretation of laws	IAB-BAMF-SOEP
The people choose their government in free elections	IAB-BAMF-SOEP/WVS
Civil rights protect the people from government oppression	IAB-BAMF-SOEP/WVS
Minorities are protected	IAB-BAMF-SOEP
Women have the same rights as men	IAB-BAMF-SOEP/WVS

We test for measurement invariance twice: Once for refugees only, and once for refugees and Germans.

Table 2: Country of Origin, Number of Respondents and Source Dataset

Country of Origin	Source Dataset	Observation (%)
Syria	IAB-BAMF-SOEP	2,229 (38.8)
Afghanistan	IAB-BAMF-SOEP	573 (10.0)
Iraq	IAB-BAMF-SOEP	594 (10.3)
Eritrea	IAB-BAMF-SOEP	302 (5.3)
Germany	WVS	2,046 (35.6)
Total		5,744 (100)

To estimate whether perceptions of democracy are comparable across languages, we rely solely on the refugee data because the WVS has only small in-country variance in languages. In the IAB-BAMF-SOEP Survey of Refugees, the target population is, first, multi-linguistic and, second, respondents are offered translated field instruments. For a first model, we group respondents by reported mother tongue (excluding languages used by very small numbers of respondents). In a second model, we use groups defined according to the language chosen by respondents to complete the questionnaire (see Table 3). Respondents could choose between German, English, Farsi/Dari, Pashto, Urdu, Arabic, and Kurmanji (Jacobsen, 2018). Due to low usage as a survey language, Pashto and Urdu are omitted from all estimations. Table 3 displays the distribution of mother tongues and the choice of survey language.

Table 3: Mother Tongue, Survey Language, Number of Respondents

Mother Tongue	Observation (%)	Survey Language (%)	Observation
Albanian	121 (3.2)	English	728 (16.5)
Arabic	2,062 (53.8)	Arabic	2,952 (67.0)
Dari/Farsi	533 (13.9)	Farsi/Dari	571 (13.0)
Kurmanji	779 (20.3)	Kurmanji	152 (3.5)
Pashto	70 (1.8)		
Somali	54 (1.4)		
Tigrinya	217 (5.7)		
Total	3,836 (100)		4403 (100)

4. Results

We started by testing for cross-national measurement invariance within the refugee population. In order to show that the latent construct actually exists in the data, we first conduct a non-grouped confirmatory factor analysis (CFA). Table 4 displays the factor loadings of a CFA within the IAB-BAMF-SOEP Survey of Refugees. As indicated by their low loadings, for two manifest variables, it is at least questionable whether they are explained by the latent construct: “The government taxes the rich and supports the poor” and “religious leaders ultimately determine the interpretation of laws”. Both factor loadings are substantially lower than the others.

Table 4: Confirmatory Factor Analyses Without Groups – Refugees Only

	Model 1	Model 2	Model 3
Variable	Loading (SE)		
The government taxes the rich and supports the poor	0.40 (0.02)	0.41 (0.02)	-
Religious leaders ultimately determine the interpretation of laws	0.21 (0.02)	-	-
The people choose their government in free elections	0.63 (0.02)	0.63 (0.02)	0.63 (0.02)
Civil rights protect the people from government oppression	0.70 (0.02)	0.71 (0.02)	0.70 (0.02)
Minorities are protected	0.80 (0.02)	0.79 (0.02)	0.80 (0.02)
Women have the same rights as men	0.64 (0.02)	0.64 (0.02)	0.63 (0.02)
Fit Indices	CFI = 0.95 RMSEA = 0.08	CFI = 0.96 RMSEA = 0.09	CFI = 0.96 RMSEA = 0.14

Therefore, in a second and third model, both variables are excluded stepwise. The fit indices indicate that model 2 has the best model fit. Therefore, the remaining five variables of the second model will be the basis for further tests of measurement invariance.

Table 5 displays the fit indices for the stepwise procedure. They indicate that configural invariance (CFI = .96) is given, whereas metric and scalar invariance are not because the CFI drops substantially (by more than 0.01). This conclusion is supported by the size of RMSEA. Individual factor loadings for each group are displayed in table A1 in the appendix. Looking at the factor loadings for all countries of origin separately reveals that for the Afghan and Syrian population, the item "the government taxes the rich and supports the poor" has a substantially lower factor loading than for other countries, indicating that these populations have a different understanding of this aspect of what constitutes a democracy.

Table 5: Fit Indices for Each Step of Measurement Invariance for Country of Origin – Refugees Only

	Full Invariance	**Partial Invariance 1**	**Partial Invariance 2**
	Comparative Fit Index / Root Mean Squared Error of Approximation		
Configural	0.96 / 0.09	0.96 / 0.09	0.96 / 0.09
Metric	0.94 / 0.09	0.94 / 0.10	0.96 / 0.09
Scalar	0.92 / 0.09	0.92 / 0.09	0.95 / 0.07

Testing for partial metric and partial scalar invariance by setting parameters for one variable free (determined by the modification indices and the expected parameter change; "civil rights protect the people from government oppression") did not improve the model. When setting additional parameters for one more item free ("the people elect the government in free elections"), the fit indices show improved model fit and less deterioration of the CFI. This might point to the conclusion that means across groups could be compared meaningfully if setting parameters for all but three items free is considered adequate. However, as we observe an absence of strict measurement invariance, we conclude that between-group comparisons regarding country of origin of refugees is likely to be problematic.

Table 6: Confirmatory Factor Analyses Without Groups – Refugees and Germans

	(1)	(2)
Variable	Loading (SE)	
The government taxes the rich and supports the poor	0.42 (0.02)	-
The people choose their government in free elections	0.69 (0.02)	0.73 (0.02)
Civil rights protect the people from government oppression	0.57 (0.02)	0.53 (0.02)
Women have the same rights as men	0.56 (0.02)	0.55 (0.02)
Fit Indices	CFI = 0.98 RMSEA = 0.07	CFI = / RMSEA = /

Additionally, we test whether conceptions of democracy are comparable between refugees and the German population. As displayed in table 1 we rely on slightly different variables, because not all variables are measured in both data sources. Again, in a first step we test whether the latent construct actually exists in the data. As table 6 indicates, the fit indices reflect reasonable model fit. Excluding the variable with the lowest factor loading as a robustness check would lead to a just identified model, thus fit indices are not estimable. We therefore decide to proceed with model 1.

The use of multi-group confirmatory factor analysis (MGCFA) reveals that for some countries, the factor loadings are quite small, indicating that configural invariance might be difficult to achieve (see table A2 in the appendix). However, the fit indices indicate good model fit. Therefore, we proceed with the measurement invariance test. Table 7 indicates that configural as well as metric invariance are achieved (even though the CFI drops by more than 0.01 it is still relatively high in absolute terms) – but not scalar invariance. In order to test for partial scalar invariance, we do not restrict the parameters for "the government taxes the rich and supports the poor". Nevertheless, the CFI does not improve substantially, and the deterioration

of the CFI remains the same. Setting an additional parameter free ("the people choose their government in free elections") does not change these results. Additionally, the RMSEA supports the conclusion that the mean comparisons are problematic.

Both findings together—that within national groups of refugees only partial measurement invariance is achieved, and between national groups of refugees and Germans only metric invariance—leads us to the following conclusion: Conceptions of democracy are most likely not comparable between refugees from different countries or between refugees and the German population. Thus, we can consider Hypothesis 1 to be confirmed.

Table 7: Fit Indices for each step of Measurement Invariance for Country of Origin – Refugees and Germans

	Full Invariance	**Partial Invariance 1**	**Partial Invariance 2**
	Comparative Fit Index / Root Mean Squared Error of Approximation		
Configural	0.99 / 0.05	0.99 / 0.05	0.99 / 0.05
Metric invariance	0.96 / 0.07	0.96 / 0.06	0.97 / 0.05
Scalar invariance	0.80 / 0.12	0.80 / 0.13	0.80 / 0.13

3.1 Cross-Linguistic Comparisons of Conceptions of Democracy

In a second step, we replicate the previous analyses. However, instead of grouping over country of origin, we use language groups. As a robustness check, we use two different strategies. First, we test for measurement invariance between mother tongues, and second, we group by the language used in the survey.

Testing the second hypothesis regarding comparability between different languages reveals a similar picture. Again, in a first step, we estimate whether the latent construct exists in all groups separately. Again, for some groups, single factor loadings are somewhat too small (see table A3 in the appendix). However, the CFI for the mother tongue indicates that the latent construct exists in all groups.

Table 8: Fit Indices for Each Step of Measurement Invariance for Mother Tongue – Refugees Only

	Full Invariance	**Partial Invariance 1**	**Partial Invariance 2**
	Comparative Fit Index / Root Mean Squared Error of Approximation		
Configural	0.96 / 0.09	0.96 / 0.09	0.96 / 0.09
Metric	0.93 / 0.10	0.93 / 0.10	0.94 / 0.10
Scalar	0.91 / 0.09	0.92 / 0.09	0.92 / 0.09

When testing for metric invariance, however, we find neither full nor partial metric invariance between different mother tongues, as shown in Table 8 (for partial invariance 1, we set parameters free for the item "minorities are protected"). The same can be seen in Table 9 for the survey language (for partial invariance 1, we set parameters free for the item "women have

the same rights as men"; factor loadings within groups are displayed in Table A4 in the appendix). Thus, all models for languages (mother tongue and survey language) indicate that strict measurement invariance is not given. Setting additional parameters free ("civil rights protect the people from government oppression" for mother tongue; "minorities are protected" for survey language) in each model does not change this conclusion although the CFIs slightly improve – the model regarding survey language in Table 9 indicates that mean comparisons based on partial measurement invariance could be valid. However, the overall picture regarding language does not indicate robustness regarding group mean comparisons as strict invariance is not given and the deterioration of the CFI for partial scalar invariance based on the mother tongue is too large. Thus, we accept Hypothesis 2, which states that perceptions of democracy are most likely not comparable across languages.

Table 9: Fit Indices for Each Step of Measurement Invariance for Survey Language – Refugees Only

	Full Invariance	**Partial Invariance 1**	**Partial Invariance 2**
	Comparative Fit Index / Root Mean Squared Error of Approximation		
Configural	0.96 / 0.09	0.96 / 0.09	0.96 / 0.09
Metric	0.94 / 0.09	0.94 / 0.10	0.95 / 0.09
Scalar	0.92 / 0.09	0.92 / 0.09	0.94 / 0.08

Thus, we accept Hypothesis 2, which states that perceptions of democracy are likely not comparable across languages.

5. Discussion

5.1 Can We Compare Perceptions of Democracy in Cross-Linguistic and Cross-National Research?

In this paper, we examine whether conceptions of democracy are comparable cross-culturally and cross-linguistically in a nationally and culturally diverse sample of refugees and asylum seekers in Germany. Adding to previous research based primarily on between-country comparisons, we show that conceptions of democracy are also problematic to compare cross-culturally or cross-linguistically within the same societal context. In light of the ongoing public debate over the comparability of cultural values between refugees and Germans, this leads us to conclude that these instruments do not allow for strictly reliable conclusions concerning respondents' democratic values because they likely are unfit to capture different underlying conceptions of democracy.

Our results support previous research showing that the democracy scales in the WVS are not adequate to compare conceptions of democracy in cross-national and cross-cultural samples

(e.g. Alemán & Woods, 2016). Furthermore, we provide new insights showing that such conceptions are also not comparable across mother tongues or survey languages.

As previous research suggests, the main reason for the incomparability between different countries of origin are the different political cultures in which respondents are brought up and socialized, which engender different concepts of democracy. Thus, the very reason why the comparison of conceptions of democracy draws so much attention is the same reason why such comparisons should be treated with caution. Caution is warranted particularly when studies fail to test for measurement invariance and assume that perceptions of democracy are in fact comparable across groups of interest. If there is no uniform conception of democracy, as is the case among refugees in Germany, it is equally misguided to treat democratic values as universally valid constructs. Our research suggests that the longstanding measure of democratic values in the WVS is not suited to measuring these values. Hence, one reason for the notorious problems comparative values studies have faced in explaining the divergent effects of economic development on the liberalization of values between Western and non-Western countries could lie in the fact that the WVS items measure only the Western liberal understanding of democracy, treating it as the only possible option.

Regarding the non-comparability of conceptions of democracy across different languages, we argue that languages evolve historically and that translations therefore sometimes incorporate different meanings. Additionally, language is embedded within a cultural frame, which is connected with different images and connotations for the same concept. Therefore, it is likely that questions are interpreted differently across languages (Bond and Yang 1982; Davidov and Beuckelaer 2010; Luna, Ringberg and Peracchio 2008). This conclusion is further supported by the finding that mean comparisons are especially problematic between mother tongues and to a lesser degree between survey languages – as some respondents have to rely on their second language to answer the questionnaire.

However, we would not argue that conceptions of democracy are not comparable per se. First, we see that although all of the tests suggest an absence of strict measurement invariance, the CFI in many cases is just below the threshold or even partial measurement invariance is achieved. This is a hint that the items might only be mildly problematic. We assume that the non-comparability in our data is due to the fact that the questions used in the WVS survey (which have served as a model for many other studies such as the employed refugee study in this paper) seem to reflect a Western understanding of liberal democracy. Therefore, we wonder whether our findings would hold when replicating this study with another, broader, definition of democracy (see e.g. Gabriel, 2020 who presents a different approach to estimate

the understanding of democracy using WVS data). Moreover, as demonstrated by previous international research, support for democracy does not necessarily reflect the democratic state of a country. Thus, support for some items may be driven by a desire for an abstract idea of democracy rather than a critical understanding of the concept of democracy, a relationship that has not been sufficiently researched. Qualitative research may shed light on this and indicate whether there might exist a broader, more universal understanding of democracy.

5.2 Outlook: Invalid Comparisons Have Political Implications

Our findings have important implications for future research on integration and for comparative values studies. As scholars before us have already emphasized, when comparing different latent constructs, these constructs need to be tested for measurement invariance in order to show that a comparison is valid. Depending on the research question at hand, invariance should be estimated for different groups.

Regarding survey quality, we would argue that tests of measurement invariance may be a useful tool when conducting pretests for surveys. Our example shows that the manifest variables from the WVS do not represent the same latent construct in the refugee data. Thorough pretests can avoid such misspecifications and might lead to the development of new and more appropriate items for comparative values research.

Additionally, papers addressing political values as a marker of integration should ensure that they base their analyses on a latent construct that actually reflects the intended subject equally in all groups under investigation. Thus, besides the implications of our study for the quality of such analyses, there is also a political dimension: If the latent constructs are not comparable and scholars find substantial differences in value conceptions (e.g., some foreign nationals show lower support for democratic values), this can create a negative narrative based on flawed analyses. Existing democratic attitude items should therefore not be used as a sole basis for conclusions about whether the consensus over democratic values in Germany is in jeopardy. Caution is imperative when talking about value consensus, national values, or presumed disruptions in these values due to migration when considering that previous studies found that value consensus is no defining feature of democracies *per se* (Schwartz & Sagie, 2000). Due to the strong political and societal implications, we propose that measurement invariance tests dealing with such delicate topics should be strict and upfront in their evaluation criteria in order to impede a normatively biased interpretation of the models.

As mentioned at the outset of this paper, the strong suit of our study lies in the comparison of migrants within a receiving country. While we conclude that conceptions of democracy are

likely not comparable between recent immigrants and the local population under examination, it would be interesting to see whether this changes over time and whether past experiences are in a sense overwritten by experiences in a new political environment.

5.3 Limitations

Some limitations of this study should be noted. First, it is unfortunate that we had to rely on two datasets to compare refugees and the German population. This might introduce some error resulting from different modes of data collection, the different institutes conducting the fieldwork, and different incentive strategies. Additionally, fieldwork for the WVS took place two years prior to the IAB-BAMF-SOEP Survey of Refugees. However, we assume that this does not bias our analyses because it is unlikely that such an important concept as democratic values changes substantially within three years. Furthermore, the cross-country comparison relies on a harmonization of scales. While the harmonization procedure was straightforward, it still would have been better if the answers had been collected using the same scale in the first place. The results of the cross-country comparisons should thus be viewed in light of the harmonization of scales. Finally, some tests for (partial) scalar invariance were just below the threshold of the CFI (0.95). From a critical stance one might argue that conceptions of democracy are therefore, in contrary to our conclusion, indeed comparable. However, in order to minimize the researcher's degree of freedom, we refrain from altering the way of interpretation aposterioi. Additionally, beyond being below our cut-off criterion, the deterioration of the CFI in those cases is beyond the recommended criteria as well. Moreover, as our results are in line with previous research on this matter (e.g. Alemán & Woods, 2016) and in line with other studies who used the same cut-off criterion (e.g. Hu & Bentler, 1999; Chen, 2007), we are confident that the used threshold does not pose a problem. Yet, we should note that freeing parameters for all variables but two (in spite of the recommendations in the literature) for the refugee only models would lead to partial measurement invariance for between country and between language comparisons (not displayed as a table).

Therefore, we conclude that different interpretations of partial measurement invariance or the use of more liberal cut off criteria may engender other conclusions regarding the comparability of conception of democracy in immigration societies, however, only if such more lenient interpretations are accompanied by very careful theoretical and contextual arguments. On a different note, we suppose that conceptions of democracy might align over time and what was determined incomparable in this article might be comparable in the future when refugees have lived in Germany longer and their German language proficiency has improved.

Although our study faced some obstacles, it clearly provides new insights for comparative value research. We strongly suggest that future cross-cultural, cross-country, and cross-linguistic comparative research on values be carried out with caution, and that it be backed up by an assessment of measurement invariance—even when the target population lives in the same country.

References

Alexander, A. C., & Welzel, C. (2011). Islam and Patriarchy: How Robust is Muslim Support for Patriarchal Values? *World Value Research, 4*(2). doi:10.2139/ssrn.2390634

Alemán, J., & Woods, D. (2016). Value Orientation From the World Values Survey: How Comparable Are They Cross-Nationally? *Comparative Political Studies, 49(8),* 1039-1067. doi: 10.1177/0010414015600458

Alvarez, A. M., & Welzel, C. (2014). Enlightening People. In R. J. Dalton & C. Welzel (Eds.), *The Civic Culture Transformed: From Allegiant to Assertive Citizens*. Cambridge: Cambridge University Press.

Ariely, G., & Davidov, E. (2011). Can we Rate Public Support for Democracy in a Comparable Way? Cross-National Equivalence of Democratic Attitudes in the World Value Survey. *Social Indicators Research, 104*(2), 271-286. doi:10.1007/s11205-010-9693-5

Ariely, G., & Davidov, E. (2012). Asessment of Measurement Equivalence with Cross-National and Longitudinal Surveys in Political Science. *European Political Science,* 11, 263-377. doi:10.1057/eps.2011.11

Banulescu-Bogdan, N., & Benton, M. (2017). *In Search of Common Values amid Large-Scale Immigrant Integration Pressures*. Retrieved from https://www.migrationpolicy.org/research/search-common-values-amid-large-scale-immigrant-integration-pressures, last visited: 02.04.2020

Behr, D., Brzoska, P., & Schoua-Glusberg, A. (2018). Linguistic and Cultura Aspects in Migrant Surveys: Introduction and Overview. In D. Behr (Ed.), *Surveying the Migrant Population: Consideration of Linguistic and Cultural Issues* (pp. 5-12). Köln: GESIS.

Belot, M. V. K., & Hatton, T. J. (2012). Immigrant Selection in the OECD. *Scandinavian Journal of Economics, 114*(4), 1105-1138. doi:10.1111/j.1467-9442.2012.01721.x

Bond, M. H., & Yang, K.-S. (1982). Ethnic Affirmation Versus Cross-Cultural Accommodation: The Variable Impact of Questionnaire Language on Chinese Bilinguals from Hong Kong. *Journal of Cross-Cultural Psychology, 13*(2), 169-185. doi:10.1177/0022002182013002003

Bratton, M., & Mattes, R. (2001). Support for Democracy in Africa: Intrinsic or Instrumental? *British Journal of Political Science, 31*(3), 447-474. doi:10.1017/S0007123401000175

Brücker, H., Rother, N., Schupp, J., von Gostomski, C. B., Böhm, A., Fendel, T., ... & Kühne, S. (2016). Forced migration, arrival in Germany, and first steps toward integration. *DIW Economic Bulletin*, *6*(48), 541-556.

Buber-Ennser, I., Kohlenberger, J., Rengs, B., Al Zalak, Z., Goujon, A., Striessnig, E., Potančoková, M., Gisser, R., Testa, M. & Lutz, W. (2016). Human capital, values, and attitudes of persons seeking refuge in Austria in 2015. *PLoS one*, 11(9). doi:10.1371/journal.pone.0163481

Bueno, O. (2012). Perception and Conception: Shaping Human Minds. *Biosemiotics, 6*(3), 323-336. doi:10.1007/s12304-013-9170-z

Canache, D. (2012). Citizens' Conceptualizations of Democracy: Structural Complexity, Substantive Content, and Political Significance. *Comparative Political Studies, 45*(9), 1132-1158. doi:10.1177/0010414011434009

Canache, D., Mondak, H. J., & Seligson, M. A. (2001). Meaning and Measurement in Cross-National Research on Satisfaction with Democracy. *Public Opinion Quartely, 65*(4), 506-528. doi:10.1086/323576

Chen, F. F. (2007). Sensitivity of Goodness of Fit Indexes to Lack of Measurement Invariance. *Structural Equation Modelling: A Multidisciplinary Journal, 14*(3), 464-504. doi:10.1080/10705510701301834

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating Goodness-of-Fit Indexes for Testing Measurement Invariance. *Structural Equation Modelling: A Multidisciplinary Journal, 9*(2), 233-255. doi:10.1207/S15328007SEM0902_5

Cho, Y. (2014). To Know Democracy Is to Love It: A Cross-National Analysis of Democratic Understanding and Political Support for Democracy. *Political Research Quarterly, 67*(3), 478-488. doi:10.1177/1065912914532721
Comanaru, R., & d'Ardenne, J. (2018). The Development of a Research Programme to Translate and Test the Personal Well-being Questions in Sylheti and Urdu. In D. Behr (Ed.), *Surveying the Migrant Population: Consideration of Linguistic and Cultural Issues* (pp. 13-28). Köln: GESIS.
Davidov, E., & Beuckelaer, A. D. (2010). How harmful are survey translations? A test with Schwartz's human values instrument. *International Journal of Public Opinion Research, 22*(4), 485-510. doi:10.1093/ijpor/edq030
Davidov, E., Meuleman, B., Cieciuch, J., Schmidt, P., & Billiet, J. (2014). Measurement equivalence in cross-national research. *Annual Review of Sociology, 40*, 55-75.
Diamond, L., & Linz, J. J. (1989). Introduction: Politics, Society, and Democracy in Latin America. In L. Diamond, J. J. Linz, & S. M. Lipset (Eds.), *Democracy in Developing Countries: Latin America* (pp. 1-58). London: Adamantine Press Limited.
Diamond, L., & Plattner, M. F. (2015). *Democracy in Decline?* . Baltimore: John Hopkins University Press.
Docquier, F., Tansel, A., & Turati, R. (2018). Do emigrants self-select along cultural traits? Evidence from the MENA countries. *FERDI Working Papers*(220).
Dustmann, C., Fasani, F., Frattini, T., Minale, L., & Schönberg, U. (2017). On the Economics and Politics of Refugee Migration. *Economic Policy, 32*(91), 497-550.
Finkel, S. E., & Smith, A. E. (2011). Civic Education, Political Discussion, and the Social Transmission of Democratic Knowledge and Values in a New Democracy: Kenya 2002. *American Journal of Political Science, 55*(2), 417-435.
Fischer, R., & Schwartz, S. (2011). Whence Differences in Value Priorities?: Individual, Cultural, or Artifactual Sources. *Journal of Cross-Cultural Psychology, 42*(7), 1127-1144. doi:10.1177/0022022110381429
Gabriel, O. W. (2020). Einstellungen zur Demokratie. In T. Faas, O. W. Gabriel, & J. Maier (Eds.), *Politikwissenschaftliche Einstellungs- und Verhaltensforschung* (pp. 230-247). Baden-Baden: Nomos.
Goerman, P., Meyers, M., & Trejo, Y. G. (2018). The Place of Expert Review in Translation and Questionnaire Evaluation for Hard-to-Count Populations in National Surveys. In D. Behr (Ed.), *Surveying the Migrant Population: Consideration of Linguistic and Cultural Issues* (pp. 29-42). Köln: GESIS.
Grube, J. W., Mayton, D. M., & Ball-Rokeach, S. J. (1994). Inducing Change in Values, Attitudes, and Behaviors: Belief System Theory and the Method of Value Self-Confrontation. *Journal of Social Issues, 50*(4), 153-173. doi:10.1111/j.1540-4560.1994.tb01202.x
Hu, L., Bentler, P. M. (1999). Cutoff Criteria for fit Indixes in Covariance Structure Analysis: Conventional Criteria versus New Alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6(1),* 1-55, doi:10.1080/10705519909540118
Inglehart, R. (2000). Globalization and Postmodern Values. *The Washington Quarterly, 23*(1), 215-228.
Inglehart, R., & Baker, W. E. (2000). Modernization, Cultural Change, and the Persistence of Traditional Values. *American Sociological Review, 65*(1), 19-51. doi:10.2307/2657288
Inglehart, R., Haerpfer, C. W., Moreno, A., Welzel, C., Kizilova, K., Diez-Medrano, J., . . . Puranen, B. (2014). *World Value Survey wave 6 (2010-2014)*.
Inglehart, R., & Norris, P. (2003). The True Clash of Civilizations. *Forein Policy*(135), 62-70. doi:10.2307/3183594
Inglehart, R., & Welzel, C. (2005). *Modernization, Cultural Change, and Democracy: The Human Development Sequence*. Cambridge: Cambridge University Press.
Inglehart, R., & Welzel, C. (2009). How Development Leads to Democracy: What We Know About Modernization. *Foreign Affairs, 88*(2), 33-48.

Jacobsen, J. (2018). Language Barriers during the Fieldwork of the IAB-BAMF-SOEP Survey of Refugees in Germany. In D. Behr (Ed.), *Surveying the Migrant Population: Consideration of Linguistic and Cultural Issues* (pp. 75-84). Köln: GESIS.

Jöreskog, K. G. (1071). Simultaneous Factor Analysis in Several Populations. *Psychometrika, 36*(4), 409-426. doi:10.1007/BF02291366

Kluckhohn, C. (1951). Values and Value-Orientations in the Theory of Action. In T. Parsons, E. A. Shils, E. C. Tolman, G. W. Allport, C. Kluckhohn, H. A. Murry, R. R. Sears, R. D. Sheldon, & S. A. Stouffer (Eds.), *Toward a General Theory of Action* (pp. 388-433). Cambridge: Harvard University Press.

Kostenko, V. V., Kuzmuchev, P. A., & Ponarin, E. D. (2016). Attitudes towards Gender Equality and Perception of Democracy in the Arab World. *Democratization, 23*(5), 862-891. doi:10.1080/13510347.2015.1039994

Kühne, S., Jacobsen, J., & Kroh, M. (2019). Sampling in Times of HighImmigration: The Survey Process of the IAB-BAMF-SOEP Survey ofRefugees. *Survey Methods: Insights from the Field*. doi:10.13094/SMIF-2019-00005

Linde, J., & Ekman, J. (2003). Satisfaction with Democracy: A Note on a Frequently used Indicator in Comparative Politics. *European Journal of Political Research, 42*(3), 391-408. doi:10.1111/1475-6765.00089

Linz, J. J., & Stepan, A. (1996). *Problems of Democratic Transition and Consolidation: Southern Europe, South America, and Post-Communist Europe*. Baltimore: John Hopkins University Press.

Luna, D., Ringberg, T., & Peracchio, L. A. (2008). One Individual, Two Identities: Frame Switching among Biculturals. *Journal of Consumer Research, 35*(2), 279-293. doi:10.1086/586914

Marsh, H. W., Hau, K., Wen, Z. (2004). In Search of Golden Rules: Comment on Hypothesis-Testing Approaches to Setting Cutoff Values for Fit Indexes and Dangers in Overgeneralizing Hu and Bentler's (1999) Findings. *Structural Equation Modeling: A Multidisciplinary Journal, 11(3),* 320-342- doi:doi.org/10.1207/s15328007sem1103_2

Marini, M. M. (2000). Social Values and Norms. *Encyclopedia of Sociology, 4*, 2828-2840.

McNeish, D., An, J., & Hancock, G. R. (2017). The Thorny Relation Between Measurement Quality and Fit Index Cutoffs in Latent Variable Models. *Journal of Personality Assessment, 100(1),* 43-52. doi:10.1080/00223891.2017.1281286

Medina, T. R., Smith, S. N., & Long, J. S. (2009). Measurement Models Matter: Implicit Assumptions and Cross-National Research. *International Journal of Public Opinion Research, 21*(3), 333-361. doi:10.1093/ijpor/edp037

Miller, A. H., Hesli, V. L., & Reisinger, W. M. (1997). Conceptions of Democracy Among Mass and Elite in Post-Soviet Societies. *British Journal of Political Science, 27*(2), 157-190. doi:10.1017/S0007123497000100

Müller-Hilmer, R., & Gagné, J. (2018). *Was verbindent, was trennt die Deutschen? - Werte und Konfliktlinien in der deutschen Wählerschaft im Jahr 2017*. Retrieved from https://www.boeckler.de/pdf/p_fofoe_report_002_2018.pdf, last visited: 02.04.2020

Muthen, B. O., & Satorra, A. (1995). Complex Sample Data in Structural Equation Modeling. *Sociological Methodology, 25*, 267-316.

Norris, P., & Inglehart, R. (2012). *Sacred and Secular: Religion and Politics World Wide*. Cambridge: Cambridge University Press.

Parsons, T. (1968). *The Structure of Social Action*. New York: Free Press.

Partridge, P. (1971). *Consent and Consensus*. New York: Praeger.

Rizzo, H., Abdel-Latif, A.-H., & Meyer, K. (2007). The Relationship Between Gender Equality and Democracy: A Comparison of Arab versus Non-Arab Musilm Societies. *Sociology, 41*(6), 1151-1170. doi:10.1177/0038038507082320

Rosseel, Y. (2019). *The lavaan tutorial*. Retrieved from: http://lavaan.ugent.be/tutorial/tutorial.pdf, last visited: 02.04.2020

Saris, W. E., Pirralha, A., & Zavala-Rojas, D. (2018). Testing the Comparability of Different Types of Social Indicators Across Groups. *Social Indicators Research, 135*(3), 927-939. doi:10.1007/s11205-016-1506-z

Schafer, J. L., & Graham, J. W. (2002). Missing Data: Our View of the State of the Art. *Psychological Methods, 7(2),* 147-177.

Schedler, A., & Sarsfield, R. (2007). Democrats with Adjectives: Linking Direct and Indirect Measures of Democratic Support. *European Journal of Political Research, 46*(5), 637-659. doi:10.1111/j.1475-6765.2007.00708.x

Schwartz, S. H., & Sagie, G. (2000). Value Consensus and Importance: A Cross-National Study. *Journal of Cross-Cultural Psychology, 31*(4), 465-497. doi:10.1177/0022022100031004003

Shin, D. C. (2007). Democratization: Perspectives From Global Citizenries. In R. J. Dalton & H.-D. Klingemann (Eds.), *The Oxford Handbook of Political Behavior* (pp. 259-282). Oxford: Oxford University Press.

Silver, B. D., & Dowley, K. M. (2000). Measuring Political Culture in Multiethnic Societies: Reaggregating the World Values Survey. *Comparative Political Studies, 33*(4), 517-550. doi:10.1177/0010414000033004004

Steenkamp, J.-B. E. M., & Baumgartner, H. (1998). Assessing Measruement Invariance in Cross-National Consumer Research. *Journal of Consumer Research, 25(1),* 78-107. doi:10.1086/209528

SVR. (2019). Andere Länder, andere Sitten? Welche kulturellen Unterschiede Flüchtlinge wahrnehmen - und wie sie damit umgehen. SVR Policy Brief, September 2019

Tausch, A., & Heshmati, A. (2003). Islamism and Gender Relations in the Muslim World as Reflected in the Recent World Value Survey Data. *Society and Economy, 38*(4), 427-453. doi:10.1556/204.2016.38.4.1

Tausch, A. (2016). Muslim Immigration Continues to Divide Europe: A Quantitative Analysis of European Social Survey Data. *Middle East Review of International Affairs*, *20*(2).

Ulbricht, T. (2018). Perceptions and Coneceptions of Democracy: Applying Thick Conecepts of Democracy to Reassess Desires for Democracy. *Comparative Political Studies, 51*(11), 1387-1440. doi:10.1177/0010414018758751

Vandenberg, R. J., & Lance, C. E. (2000). A Review and Synthesis of the Measurement Invariance Literature: Suggestion, Practices, and Recommandations for Organizational Research. *Origanizational Research Methods, 3*(1), 4-70. doi:10.1177/109442810031002

Vlas, N., & Gherghina, S. (2012). Where does Religion meet Democracy? A Comparative Analysis of Attitudes in Europe. *International Political Science Review, 33*(3), 336-351. doi:10.1177/0192512111415983

Wan, C., Chiu, C.-Y., Tam, K.-P., Lee, S.-L., & Lau, I. Y.-M. (2007). Perceived cultural importance and actual self-importance of values in cultural identification. *Journal of Personality and Social Psychology, 92*(2), 337-354. doi:10.1037/0022-3514.92.2.337

Welzel, C. (2013). *Freedom Rising*. Cambridge: Cambridge University Press.

Wimmer, A., & Soehl, T. (2014). Blocked Acculturation: Cultural Heterodoxy among Europe's Immigrants. *American Journal of Sociology, 120*(1), 146-186. doi:10.1086/677207

Zavala-Rojas, D. (2018). Exploring Language Effects in Crosscultural Survey Research: Does the Language of Administration Affect Answers About Politics? *methods, data, analyses, 12*(1), 127-150. doi:10.12758/mda.2017.16

Zavala-Rojas, D., & Saris, W. E. (2017). Measurement Invariance in Multilingual Survey Research: The Role of the Language of the Questionnaire. *Social Indicators Research, 140*(2485-510). doi:10.1007/s11205-017-1787-x

Appendix

Table A1: Grouped Confirmatory Factor Analysis by Country of Origin – Refugees Only

Variable	Loading (SE)	
	Syria	**Iraq**
The government taxes the rich and supports the poor	0.23 (0.03)	0.52 (0.04)
The people choose their government in free elections	0.74 (0.02)	0.74 (0.04)
Civil rights protect the people from government oppression	0.66 (0.02)	0.80 (0.04)
Minorities are protected	0.75 (0.02)	0.87 (0.04)
Women have the same rights as men	0.52 (0.02)	0.74 (0.04)
	Afghanistan	**Eritrea**
The government taxes the rich and supports the poor	0.51 (0.05)	0.39 (0.08)
The people choose their government in free elections	0.43 (0.05)	0.66 (0.08)
Civil rights protect the people from government oppression	0.62 (0.06)	0.47 (0.08)
Minorities are protected	0.77 (0.06)	0.51 (0.08)
Women have the same rights as men	0.71 (0.05)	0.76 (0.08)
Fit Indices		CFI = 0.96 RMSEA = 0.09

Table A2: Grouped Confirmatory Factor Analysis by Country of Origin – Refugees and Germans

Variable	Loading (SE)		
	Syria	**Iraq**	**Germany**
The government taxes the rich and supports the poor	0.26 (0.03)	0.50 (0.05)	0.35 (0.03)
The people choose their government in free elections	0.86 (0.03)	0.81 (0.04)	0.64 (0.03)
Civil rights protect the people from government oppression	0.58 (0.03)	0.78 (0.04)	0.47 (0.03)
Women have the same rights as men	0.46 (0.03)	0.71 (0.04)	0.57 (0.03)
	Afghanistan	**Eritrea**	
The government taxes the rich and supports the poor	0.53 (0.06)	0.26 (0.08)	
The people choose their government in free elections	0.61 (0.06)	0.73 (0.08)	
Civil rights protect the people from government oppression	0.56 (0.06)	0.36 (0.08)	
Women have the same rights as men	0.58 (0.06)	0.92 (0.09)	
Fit Indices	CFI = 0.99 RMSEA = 0.05		

Table A3: Grouped Confirmatory Factor Analysis by Mother Tongue – Refugees Only

Variable	Loading (SE)			
	Albanian	**Arabic**	**Pashto**	**Tigrinya**
The government taxes the rich and supports the poor	0.20 (0.11)	0.32 (0.02)	0.30 (0.14)	0.51 (0.10)
The people choose their government in free elections	0.70 (0.10)	0.81 (0.02)	0.39 (0.14)	0.49 (0.11)
Civil rights protect the people from government oppression	0.76 (0.10)	0.73 (0.02)	0.95 (0.13)	0.63 (0.11)
Minorities are protected	0.73 (0.10)	0.77 (0.02)	0.76 (0.16)	0.63 (0.11)
Women have the same rights as men	0.69 (0.10)	0.56 (0.02)	0.54 (0.14)	0.41 (0.10)
	Dari/Farsi	**Kurmanji**	**Somali**	
The government taxes the rich and supports the poor	0.49 (0.05)	0.38 (0.04)	0.48 (0.17)	
The people choose their government in free elections	0.45 (0.06)	0.58 (0.04)	0.69 (0.15)	
Civil rights protect the people from government oppression	0.57 (0.06)	0.63 (0.04)	0.44 (0.16)	
Minorities are protected	0.80 (0.06)	0.79 (0.05)	0.65 (0.16)	
Women have the same rights as men	0.74 (0.05)	0.61 (0.04)	1.00 (0.12)	
Fit Indices	CFI = 0.96 RMSEA = 0.09			

Table A4: Grouped Confirmatory Factor Analysis by Survey Language – Refugees Only

Variable	Loading (SE)	
	English	**Arabic**
The government taxes the rich and supports the poor	0.37 (0.05)	0.30 (0.02)
The people choose their government in free elections	0.65 (0.05)	0.79 (0.02)
Civil rights protect the people from government oppression	0.70 (0.05)	0.69 (0.02)
Minorities are protected	0.72 (0.05)	0.79 (0.02)
Women have the same rights as men	0.60 (0.05)	0.61 (0.02)
	Farsi/Dari	**Kurmanji**
The government taxes the rich and supports the poor	0.47 (0.05)	0.64 (0.10)
The people choose their government in free elections	0.44 (0.05)	0.36 (0.11)
Civil rights protect the people from government oppression	0.59 (0.06)	0.86 (0.10)
Minorities are protected	0.79 (0.06)	0.85 (0.10)
Women have the same rights as men	0.72 (0.05)	0.43 (0.10)
Fit Indices		CFI = 0.96 RMSEA = 0.09

7 Article 4: Institutional Aspects of Credential Recognition

Preface

So far, three steps of a survey process have been analyzed: sampling, response, and validation of concepts. The last paper focuses on building hypotheses and subjects common assumptions about the economic integration of refugees in Germany to critical examination. As the paper reveals, most previous work dealing with economic integration, first, emphasized labor market access, and second, focused on individual behavior. There are two reasons why there needs to be a change in how economic integration is analyzed:

1) Labor market access might be the ultimate goal of economic integration; however, it does not describe the economic integration process itself.
2) Individual behavior always takes place in a setting of institutions that can restrict behavior. Therefore, institutions need to be incorporated into analyses.

Regarding the economic integration of refugees, I considered three institutions to be especially important to look at: recognition of degrees, integration classes, and residence status. I chose these three and not other institutions because the recognition of degrees is an institutional process that is designed to prevent loss in value of human capital, a commonly known problem triggered by migration. Therefore, the recognition of foreign credentials is an important marker of economic integration as it mirrors an investment in the future. Integration classes are important as well, because they are the key institution provided by host societies to assure that refugees are able to navigate and integrate into the new society. And refugee status by definition is important because it is the legal structure that defines and restricts refugees in their scope of action. Arguably, other institutions such as family reunification and educational attainment could have been analyzed as well. But, for the purpose of explaining economic integration, these are only secondary. As previous research has shown, the main impact of family reunification is on well-being. Second, access to educational institutions can be a factor in economic integration, but this is mainly true for young refugees. Therefore, education is more useful to examine as an institution when dealing with refugee youth.

An Investment in the Future: Institutional Aspects of Credential Recognition of Refugees in Germany

This article is published and should be cited as follows:
Jacobsen, Jannes (2019) An Investment in the Future: Institutional Aspects of Credential Recognition of Refugees in Germany. *Journal of Refugee Studies,* online first. DOI 10.1093/jrs/fez094.

Abstract
Adding to the rich literature on the economic integration of refugees, this paper extends the scope toward the role of institutions by focusing on the transfer of human capital by means of credential recognition.
The 2012 Federal Act of Recognition in Germany is a new institution that provides the possibility to study the transfer of human capital in-depth. I argue that analyzing the decision for recognition of credentials is an important aspect of economic integration because it mirrors an investment in future labor market access. I hypothesize that institutions, such as integration and language classes and the refugee status by admission, are the key for explaining who tries to obtain official recognition for their credentials. In order to test my hypothesis, I employ panel data from the IAB-BAMF-SOEP Survey of Refugees in Germany. Results show that both a secure residence title and participation in language/integration classes lead to a higher likelihood of applying for the recognition of degrees.

1. Introduction

Following the unprecedented influx of asylum seekers into several European countries through Greece and the Balkans in 2015 and subsequently to North America, policy makers are facing an urgent need to improve these people's prospects for integration. For example, in 2016, around 90 percent of refugees in Germany reported their intention to stay for a longer period—in some cases forever[1]—and the ongoing crisis situation in 2018 and 2019 in countries like Somalia, Eritrea and Syria strongly suggests that many will indeed remain. In other countries, such as Canada, refugees are likely to become citizens and thus a prospect of integration is even more vital. This brings the issue of economic integration to the fore as one of the most urgent issues facing receiving countries.

Often, researchers and policy makers identify labour-market entry as the core of successful economic integration, which subsequently provides refugees with resources and possibly relieves social welfare. Therefore, most articles addressing the economic integration of refugees focus on labour-market access through the lens of employment probabilities either within migrant cohorts or between locals and migrants (Bakker, Dagevos, & Engbersen, 2017; Dumont, Liebig, Peschner, Tanay, & Xenogiani, 2016; Jacobsen, Kroh, Legewie, & Salikutluk, 2018). However, such studies neglect the fact that, even before migrants enter the labour market, they need some signal (Spence, 1978) that their job qualification is transferable to the new society. Thus, in order to understand the labour market access of migrants, institutions that determine the labour market access need to be the subject of analyses. Institutions are defined as a (formal) structure, such as laws or commonly accepted norms and rituals, that governs societal life (Hall, 2010, p. 204; Knight, 1992). One crucial institution regarding economic integration is the recognition of foreign credentials. Increasingly, countries provide the opportunity of credential recognition in order to facilitate labour market access (Australian Government - Department of Education, 2019; BMBF, 2017; Bundesministerium Europa Integration und Äußeres, 2019; Government of Canada, 2019; Swedish Council for Higher Education, 2019).

Studying economic integration through the intention of credential recognition is a good additional proxy of economic integration because it grasps the effort to transfer human capital to the local labour market by the migrant. This has not been well studied within research literature yet.

[1] Own calculation of the Data of the IAB-BAMF-SOEP Survey of Refugees.

Besides serving as an additional proxy for economic integration, I argue that institutions can also explain economic integration as an explanatory variable themselves (Bakker, Degevos, & Engbersen, 2013; Borjas, 1991; Chiswick, Lee, & Miller, 2006). In the case of refugees, it is especially important to analyse the legal status and the participation in language/integration classes (Ager & Strang, 2008). For example, the legal status determines whether an individual, legally speaking, is allowed to work and how long he or she can stay (Bakker et al., 2013; BAMF, 2016a), while integration/language classes provide individuals with a network, language proficiency, and institutional knowledge (BAMF, 2016b; Hoehne & Michalowski, 2016). Hence, refugees become more familiar with how institutions work and, therefore, how they can make use of them.

With this design, I add to the existing literature on economic integration of refugees, but switch the focus towards the role of institutions, namely the recognition of foreign credentials, thus using an additional indicator for economic integration. My main research question is how the legal status and the participation in language/integration classes affect the decision of degree recognition.

I use Germany as a case study in order to answer this research question. Germany has a labour market, largely based on certified skill. Thus, credential recognition is a crucial aspect of economic integration. The German recognition act was introduced in 2012 and provides the opportunity for people with foreign credentials to get a recognition. Recognition is provided for all professions, including those that are not regulated.

I employ data from the German Socio-economic Panel (SOEP) (Göbel et al., 2019). Since 1984, the SOEP has made a yearly panel study of private German households. Starting in 2016, a panel study of refugees was implemented—the IAB-BAMF-SOEP Survey of Refugees in Germany (see Brücker, Rother, & Schupp, 2016, 2017; Kroh, Kühne, Jacobsen, Siegert, & Siegers, 2017). Based on a set of retrospective questions, I construct panel data reaching back from January 2013 until March 2018 and employ linear-probability-regression analysis. In order to avoid endogeneity, I use a coarsened exact matching strategy and estimate an analytical weight that equalizes the distribution of confounding factors between dependent and independent variables.

Choosing only refugees as the target population to study institutional factors of credential recognition has one big advantage: in contrast to most other migrants, their residence permit does not depend on economic factors (employment, credentials, etc.), but relies on persecution and thus is exogenous to economic integration. For the same reason, I assume that

the participation in integration classes is exogenous to credential recognition as well. Therefore, I can exploit the variation in the refugee status by admission and the attendance of integration classes to estimate their effect on the intention to obtain credential recognition. More abstract, this allows the study of whether the provision of opportunities and knowledge leads to an increasing effort in economic integration.

Findings suggest that a secure residence permit and participation in language/integration courses increase the chance of applying for degree recognition.

2. Institutions as Behavioural Restrictions and Incentives

There are different approaches to explain the circumstances under which refugees[2] access foreign labour markets. This article focuses on institutions. Institutions are defined as a (formal) structure, such as laws or commonly accepted norms and rituals, that governs societal life (Hall, 2010, p. 204; Knight, 1992). It is important to mention that I do not overlook the fact that, for example, individual motivation plays a substantial part in integration and should be considered in the analysis of integration processes (see Berry, 1997; Chin & Cortes, 2015; Pietkaza-Nykaza, 2014). But, as Ager and Strang indicate, people are only able to put their personal traits and talents to use when the institutional framework allows them to do so (Ager & Strang, 2008). Thus, when talking about the micro level, it is important to look at institutions as the larger framework and how they restrict or incentivize specific behaviours. When arguing that integration entails specific forms of consumption, social interaction, employment, or access to multiethnic networks (see also Cheung & Phillimore, 2014, p. 533; Eisnecker, 2019; Esser, 2006; Phillimore & Goodson, 2006), it should be kept in mind that institutions can restrict individuals in their decision on how to participate in the society (see also Portes & Rumbaut, 2001; Pries, 2016). Thus, in order to understand integration, institutions must be incorporated into theory and analyses.

Prominent in this regard is the work of Ager and Strang (Ager & Strang, 2008). Based on qualitative interviews with refugees, Ager and Strang developed a paradigm of integration in which rights and citizenship are the *foundation* to subordinate dimensions of integration (Ager & Strang, 2008). After the foundation is set by means of providing rights, facilitators such as cultural knowledge and stability come into play. Facilitators remove barriers that stand in the way of integration such as language issues. Social connections as a third layer are important to mention as well. Social ties to peers, other communities and to the 'structure of the state' (Ager & Strang, 2008, p. 178) provide opportunities and are supposed to improve the quality of life in

[2] The term "refugee" is used differently in the literature. I use the term as a synonym for all forms of humanitarian protection.

general and provide access to relevant services. Markers and means of integration are then specific key areas of integration, such as health, employment or education. Access to these keys areas can be a proxy for successful integration.

Applying this conceptual framework to the case of Germany, I identify two institutions that are crucial in order to understand economic integration: the refugee status by admission as the foundation for getting access to societal domains and the participation in integration classes as a facilitator because it removes barriers such as cultural knowledge and language issues. While the original framework by Ager and Strang provides a comprehensive framework of integration in general, including sub-domains such as social and economic integration, I will focus on economic integration only. Usually, in the literature, economic integration is measured by means of employment. However, I argue that this definition lacks off a perspective that focuses on the preconditions for labour-market access. Thus, economic integration measured by the employment status only, misses this important aspect. Therefore, in order to understand economic integration, I argue that the preconditions need to be analysed as well. Thus, looking at the individual decision to apply for recognition widens the view on economic integration. For instance, if self-selection takes place in this step, disparity in actual labour-market placement can be explained.

2.1 The Federal Act of Recognition as an Institutional Framework

A core problem caused by migration is that human capital can lose its value (Chiswick & Miller, 1992, 2009; Friedberg, 2000). Human capital includes certified as well as informal skills of an individual (Becker, 1993; Sicherman, 1991). With regard to migration, certified skills such as vocational and academic degrees are especially affected by value loss through immigration. Usually, proof and utility of skills are linked to national regulations and particularities (i.e. regulated professions or language). Therefore, gatekeepers (those who decide about employment) on a local labour market have difficulties in identifying the value of foreign credentials (Imdorf, 2010; Seibert & Solga, 2005). According to the signalling theory (Spence, 1973, 1978), credentials provide a signal about their economic utility. If a credential, for instance, is not provided in the local language, or was granted in a country without a high educational reputation, the signal of economic utility can be mitigated. Therefore, migrants with foreign certified but unrecognized credentials are especially at risk of suffering from unemployment or underemployment–meaning labour below qualification (Brussig, Dittmer, & Knuth, 2009; Krahn, Derwing, Mulder, & Wilkinson, 2000). Current research indicates that underemployment has a negative impact on labour-market trajectories (Verbruggen, Emmerik, Gils, Meng, & Grip, 2015). This is especially true for countries where the first labour-market

entrance is usually regulated by certified skills (Blossfeld & Mayer, 1988). Applying for credential recognition is thus a good proxy to study the steps towards and the process of economic integration.

In order to facilitate such recognition, the German government successively implemented the Federal Act of Recognition (in German: *Anerkennungsgesetz des Bundes*) in 2012 and 2014. Its implementation assumed that an official statement by proper authorities on foreign degrees can upgrade their value in the local labour market. The Federal Act of Recognition is a supplement to existing legal regulations (i.e. *Berufsqualifikationsfeststellungsgesetz*; BQfG). It provides immigrants with the opportunity to obtain proof that their educational degree matches the German counterpart. The Federal Ministry of Education and Research (BMBF) provides a webpage (www.anerkennung-in-deutschland.de) where potential candidates can search for their profession and see whether recognition is a precondition for employment in that sector (e.g. teaching) or whether it is not necessary but recommended (e.g. mechatronics engineer).[3] The decision on whether recognition is obligatory or not in order to work in a specific profession depends on whether the job is legally regulated. The website is available in English, Spanish, French, Italian, Romanian, Polish, Turkish, Greek, Russian and Arabic. In order to apply for recognition, different steps are necessary:

1) Potential candidates use a search engine to find their profession on www.anerkennung-in-deutschland.de.
2) If their job is eligible for recognition, candidates make an appointment with a nearby office. These offices are not necessarily federal or local authorities. The organization providing recognition depends on whom, legally speaking, has the sovereignty over the regulation of each specific profession. If not the government, these are often chambers of commerce, chambers of handicrafts or other professional associations.
3) After the meeting, an equivalence assessment is carried out. There are three possible results:
    a. equivalent
    b. partially equivalent, including requirements in order to reach full equivalence
    c. not equivalent

[3] Academic degrees are a special case leading to an unregulated (academic) profession, such as researchers. In this case, the Central Office for Foreign Education (ZAB) provides an assessment of the academic degree. This assessment is then supplementary to the certificate. Further, school certificates can be recognized, but are not included with the recognition act. Usually, institutions like universities or vocational schools that follow up after school carry out an assessment themselves.

The process itself is supposed to take about three months. However, it varies depending upon the corresponding documents, besides the certificate, that are needed. Furthermore, candidates must pay a fee that varies depending on the profession and where the recognition takes place. Including the translation of the documents, fees can amount up to 600€- independently of the result (BMBF, 2019). The German government provides some support mechanisms. For example, applicants with a low income are entitled to a grant-in-aid.

Although scholars in educational and migration studies historically focus on the transferability of human capital between different countries (Chiswick & Miller, 2009) and the Federal Act of Recognition was implemented some time ago, little is known about who actually takes advantage of the degree recognition. Policy evaluations by the German government show that, in 2017, most applications came from Syrians (~ 3,000), people from Bosnia and Hercegovina (~ 2,400) and Serbs (~ 2,000). The most common professions were nurses/care-givers and medical doctors. Only 2 per cent of all applications in 2017 were rejected, around 9 per cent were partially recognized, 28 percent of recognitions included requirements for further training, and 60 percent were fully recognized right away (Schmitz, 2018). Considering these high numbers of (partial) recognition, this indicates that the intention to apply for recognition must be closely correlated with a positive outcome.

Regarding refugees, around one-third of those who hold a certified degree (~ 25 per cent of the adult population) have already applied for recognition (Brücker et al., 2017). This relatively low application rate is probably because refugees, when leaving their country, do not know where they will end up and therefore do not know what they need in order to access the labour market. Additionally, it is questionable whether refugees, when leaving their home country at short notice, think about taking proof of their credentials.

There is some research showing that the recognition of degrees in Germany facilitates labour-market access of labour migrants (Brücker, Liebau, Romiti, & Vallizadeh, 2014; Kogan, 2012). But, to the best of my knowledge there is no analysis addressing the role of institutions as a selection factor such as residence permits. Despite the fact that, self-selection in many cases points towards institutional hurdles that hamper access for certain groups. Moreover, to date, there has been no thorough analysis dealing with refugees in particular. However, this is necessary because refugees differ substantially from other migrants. First, refugees, when leaving their country of origin, do not know where they will end up eventually. Thus, they cannot prepare for national specificities of labour markets. Second, when fleeing one's country, it is likely that necessary documents are left behind and, third, in contrast to the situation for most other migrants, a refugee status is not linked to labour-market access.

An important institution in order to understand economic-integration trajectories is the provision of rights. Rights granted by means of citizenship or a residence permit define the access to local resources.

2.2 The Legal Status of Refugees as an Institutional Framework

There is a vast literature on how rights granted for migrants and subsequently a prospect of remaining in the country affect their integration trajectories (Ager & Strang, 2008; Bellamy, 2008; Bevelander, 2011, p. 42; Favell, 1998; Hainmueller & Hangartner, 2013; Hainmueller, Hangartner, & Pietrantuono, 2015, 2017; Pietrantuono, 2016). Usually, two hypotheses are discussed. On the one hand, scholars argue that immigrants should be granted residence permits or citizenship quickly and comprehensively because rights are the core resource for rapid integration (Hainmueller et al., 2015, 2017). On the other hand, some argue that comprehensive access to rights will hamper integration because migrants lose the incentive to invest in a future in the host country as soon as their status is secure (Koopmans, 2009).

Empirically speaking, most research points towards the conclusion that there is a positive correlation between granting rights comprehensively and integration trajectories (Ager & Strang, 2008; Bevelander, 2011; Hainmueller et al., 2015, 2017; Pietrantuono, 2016). Therefore, citizenship is usually seen as a prerequisite to integration (Favell, 1998) because it provides migrants with the same 'opportunities as anybody else in the country' (Ager & Strang, 2008, p. 176). Ager and Strang argue that having rights to participate in the society does not only provides a resource for integration, but also creates a sense of belonging (Ager & Strang, 2008; Delanty, Jones, & Wodak, 2008; Jones & Krzyzanowski, 2008). Based on qualitative interviews, Ager and Strang developed a scheme of integration in which rights and citizenship of migrants are the foundation of integration (Ager & Strang, 2008). In addition, the quantitative literature points to this conclusion. For example, research based on a natural experiment in Switzerland by the Immigration Policy Lab (Hainmueller & Hangartner, 2013) indicates that citizenship improves (i) political integration (Hainmueller et al., 2015), (ii) the intention to stay in the host country (Hainmueller et al., 2017), and (iii) labor market participation (Pietrantuono, 2016). Work by (Bevelander, 2011) indicates that a different refugee status by admission can systematically explain labour-market participation. A causal analysis for Germany and the current cohort of refugees is unfortunately missing.

In sum, the literature on citizenship and naturalization as well as the scarce research on refugee status and rights suggests, that comprehensive legal access to societal domains in the host country fosters integration. Granting rights promotes a sense of belonging and creates an incentive for integration.

In the following, I will provide an overlook on different regulations regarding refugees in Germany.

2.3 Refugee Protection in Germany

With regard to humanitarian protection in Germany, there are four different forms of protection that need to be considered: refugee status, asylum status, subsidiary protection, and suspension of deportation/toleration (for Germany see BAMF, 2016a; for a historic classification see Scheinman, 1983; UNHCR, 1966, p. 14). These four groups receive different residence permits:

(1) Individuals defined as refugees under the Geneva Convention on Refugees and those with asylum status receive a **three-year** residence permit. They are allowed to work. Permanent residency may be granted after five years if language and employment criteria are fulfilled. Participation in integration classes is allowed.
(2) Individuals with subsidiary protection receive a **one-year** residence permit, which can be extended. Permanent residency may be granted after five years if language and employment criteria are fulfilled. They are allowed to work. Participation in integration classes is allowed.
(3) Those with suspension of deportation or toleration do not receive a residence permit. However, they are not deported until the reason for a suspension invalidates (e.g. health condition improves, coming of age, missing passport resurfaces). Access to the labour market is granted on a case-by-case basis and for only one year at a time. Participation in integration classes is allowed.
(4) Individuals who are in the application process receive a temporary residence permit for the period in which they are awaiting the decision on their legal status. Those who are waiting for a first decision are not permitted to work or enter vocational training. As of 2014, they are only permitted to work (but not to enter self-employment or vocational training) if the application process takes longer than three months (Ohliger & Brands, 2016). Participation in integration classes is allowed if the probability of recognition as a refugee is higher than 50 per cent.

If someone does not fit into any of the first three categories of protection listed above, the government may also grant a residence permit for 'humanitarian reasons, international law, or on political grounds' (see §23 Residence Act). However, this status may be revoked at any time if criteria are not fulfilled anymore.

In order to mirror the debate about giving incentives for integration by means of residence status, I distinguish between two groups of refugees in Germany: those with 'secure' and 'not secure' residence permits. A permit is 'not secure' when it is either preliminary, grants a one-year stay or less (subsidiary protection, suspension of deportation), or can be revoked at any time (residence permits as defined in §§22, 23, 25 Section 3, 4 or 5 Residence Act). All other permits are considered 'secure' (see also table A1 in the appendix).

As mentioned above, the current research suggests that individuals rather invest in integration when they have a long prospect of staying. Therefore, I hypothesize that a 'secure' status provides an opportunity structure by means of the prospect that migrants can stay in Germany for a longer period. Thus, making an effort towards integration is worthwhile.

> *H1: Refugees with a secure residence permit are more likely to seek recognition of their foreign credentials.*

2.4 Integration Classes as an Institutional Framework

So far, I have argued that especially the application for credential recognition is a proxy of economic integration and the decision for degree recognition is shaped by the residence permit. However, migrants need to have knowledge about degree recognition in order to understand its impact. So-called integration and language classes are thus the third valuable institution to consider when analysing economic integration. From the perspective of migration research and with regard to the aforementioned integration paradigm by Ager and Strang, such knowledge is a 'facilitator' for participation in the society (Ager & Strang, 2008, p. 170). With regard to (Hale, 2000, p. 276), Ager and Strang argue that social participation in the local society is the key to integration. They directly address 'language and cultural knowledge' as a barrier for this to happen (Ager & Strang, 2008, p. 181) because knowing customs and the language is an important precondition for participation. Research on integration classes indicates that they are able to reduce such barriers, not only because customs are introduced, but also because learning the local language is part of the curricula (Hoehne & Michalowski, 2016; Schuller, 2011; Tubergen, 2010; Weiermair, 1976).

Typically, asylum seekers who migrate to Germany are offered integration classes. These classes provide an initial orientation to the new environment and involve two phases: 600 hours of language training and 100 hours of introduction to daily life in Germany. The participants learn important vocabulary as well as norms, values and customs that are of relevance. Information about access to the job market is provided as well as support for building networks. Further, general knowledge about how to use public transport, become

involved in clubs and participate in leisure activities is shared. Thus, these classes offer an introduction into the German institutional framework. As the Federal Office for Migration and Refugees (BAMF) sometimes hires third parties to organize these classes, the specific content and quality of a given class are unknown. I argue that such classes provide immigrants with at least three things:

1) German-language abilities;
2) institutional knowledge; and
3) a network including individuals who are not necessarily refugees.

With regard to the current research that identifies knowledge and language as a facilitator of integration, I hypothesize:

> *H2: Refugees who attend integration and German classes are more likely to seek the recognition of credentials than people who have not attended such classes.*

3. Method and Data

I work with the German SOEP (Göbel et al., 2019). In 2016, it incorporated a boost sample of refugees–the IAB-BAMF-SOEP Survey of Refugees in Germany. The IAB-BAMF-SOEP Survey of Refugees is a yearly conducted panel study of refugees residing in Germany. The mode of data collection is computer-assisted personal interviewing. The sample was drawn randomly from the German Central Register of Foreigners (AZR) and allows inference on refugees and asylum seekers who migrated to Germany during the years 2013-2016. Most participants come from Syria, Afghanistan, Iraq and Eritrea. The minimum age required for participation is 18 years at the time of the interview. The parents provide information on younger household members. Unaccompanied minors are not part of this survey. A disproportionate sampling design with regard to gender, age, country of origin and asylum status was employed in order to ensure a minimum sample size for sub-groups. Furthermore, the sample consists of several sampling tranches because, during sampling in early 2016, the inflow of refugees was still high and therefore the sampling frame was updated constantly. Design and non-response weighting procedures ensure generalizability. The IAB-BAMF-SOEP Survey of Refugees is the only general random sample of adult refugees in Germany (Jacobsen et al., 2019; Kroh et al., 2017). I make use of the first and second waves. Using both waves, the panel includes 7,294 distinct adult individuals and allows generalizability about persons older than 18 years who migrated to Germany between January 2013 and December 2016 and who appeared in the Central Register of Foreigners as refugees or asylum seekers (AZR; (Gostomski & Pupeter, 2008). Hence, the peak period of the recent refugee influx to Germany is covered.

I restrict the sample (see table 1) to individuals between the ages of 18 and 55 in order to capture only those who have, with regard to their age, a realistic chance of placing themselves in the labour market. I also exclude those who, at the time of the interview, had been in Germany for less than three months. After this initial waiting period, even asylum claimants are allowed to work. As I am specifically interested in the effect of recognized educational degrees, I further exclude those who do not hold a vocational or academic degree. The final working sample consists of 1,398 individuals (20 per cent of the gross sample) who are all eligible for a recognition of their credentials.

Table 1: Sample Restrictions

Sample Restrictions	Excludes	Total
None	-	7,294
Of working Age (18-55) by time of migration	-600	6,694
Vocational/Academic Degree	-5,289	1,405
Time of Immigration > 3 months	-7	1,398
All	5,896	1,398

As a result of these restrictions, the working sample is selective. The article at hand is able to make statements about refugees who sought asylum in Germany between 2013 and 2016, who are between 18 and 55 years old, and who have a foreign but formal vocational or academic degree. This population should represent around one-third of the current adult population of refugees in Germany (Brücker et al., 2017).

3.1 Constructing Panel Data

Although the IAB-BAMF-SOEP Survey of Refugees only provides panel data for 2016 and 2017, some information about earlier periods is available. For example, respondents provide information about the point in time at which they received humanitarian protection, participated in integration classes and applied for recognition. Therefore, using this information, I am able to construct a dataset with a panel structure that goes beyond the initial panel. This approach is superior to using only the two waves because it minimizes information loss, since many respondents already went into the recognition process before the panel started in 2016. Therefore, I would not make use of the full potential of the data if I had only measured changes between 2016 and 2017. The final panel provides monthly observations between January 2013 and March 2018. Because not all respondents from the first wave participated in the second wave again, for some respondents, I only have retrospective data until 2016. Furthermore, I only use information about the time after migration to Germany.

3.2 Operationalization

The dependent variable indicates whether a person has applied for recognition of a formal educational/academic degree. I do not rely on administrative data, but on self-reports with regard to the recognition procedure.

The two key independent variables measure whether the respondent receives a secure status and whether he or she participated in an integration or language class. Both rely on self-reports. The treatment period is defined as the month and all subsequent time points after a secure status is granted or the respondent started to participate in an integration/language class. I estimate a linear-probability model with individual and time fixed-effects with robust standard errors. Because the time frame of this panel covers each month of 2013 through 2017 and the beginning of 2018, time-constant heterogeneity is not an issue.

In order to control for confounding factors that were not measured retrospectively, I employ a coarsened exact matching and use the matching weight as an analytical weight in the estimations. Therefore, after weighting, the distribution of confounding factors is equal in treatment (secure status/participation in courses) and control group (no secure status/no participation in courses) (Imbens & Wooldridge, 2009), thus avoiding bias due to self-selection.

The procedure of matching followed a set of rules. First, based on theoretical considerations, possible confounding factors are identified (Table A2 in the Appendix). Second, only those that differ between the treatment and control groups (estimated by way of logistic-regression analysis; see Table A3 in the Appendix) are then used in the matching procedure (see Table 2).

Table 2: Documentation of Variables used in the Matching Process over "Secure Status" and "Integration/Language class"

Secure Status	Integration/Language Class
Motivation coming to Germany. 1) Because of Labor Market or educational system 2) other	Years of education 1) 1-9 years 2) 10-13 years 3) 13+ years
Years of education 1) 1-9 years 2) 10-13 years 3) 13+ years	Country of Origin with good "prospect of remaining in the country[1]"? 1) yes 2) no
German proficiency 1) good 2) medium 3) bad	German proficiency 1) good 2) medium 3) bad
Current age in terciles	
Mother tongue available on anerkennung-in-deutschland.de? 1) Yes 2) no	

[1] Note: The Federal Office for Migration and Refugees classifies asylum seekers regarding their prospect of obtaining asylum or refugee protection. Currently, those chances are high for respondents Syria, Eritrea, Iraq, Iran, and Somalia. This classification considers how many people from a respective country are provided with asylum. If this rate is higher than 50%, the chances are officially declared as "good". This classification serves as a reference of whether certain people get early access to integration classes for the time of their asylum application.

The matching procedure makes use of a coarsened exact matching. The advantage of a coarsened exact matching is that only those cases that actually share the same characteristics are matched. For a coarsened exact matching, a strata variable reflects possible combinations of different confounding factors. For example, a strata variable for two confounding factors with each two characteristics would have four expressions (2 x 2). Cases are only matched if they have the same expression on the strata variable but differ on the variable that defines the treatment and control groups. However, using a coarsened exact matching, one needs to avoid dismissing too many unmatchable cases. As shown in table 3, this is not the case in this analysis. Additionally, the Multivariate L1 distance, in both matching procedures, drops significantly before and after matching (from 0.267 for matching over a secure status and 0.268 for matching over participation in integration classes to 0.00 in both cases).

Table 3: Properties of Coarsened Exact Matching

	Secure Status		Integration/Language Class	
	Secure	Not secure	Not attended	Attended
All Cases	680	718	382	1,016
Matched Cases	584	651	381	1,001
Unmatched Cases	96	67	1	15

Collection of information on the variables used for the matching procedure took place during the first wave.

The weighted (ω) linear-probability regression with weighted individual (i), time (t)-fixed effects (d) denotes a function of

$$\Pr(Y) = \alpha + d_t + X\beta w_i + \varepsilon_{it} \quad (1)$$

Where Pr ($\bar{Y}_{it}$) = probability for recognition, α = constant, x = variance covariance matrix for independent variable(s) and controls, ε = residuals and the weighted-effect size denotes a function of

$$\beta w_i = (X'WX)^{-1}X'WY \quad (2)$$

(DuMouchel and Duncan 1983).

In order to avoid an artificially inflated total and, therefore, an artificially efficient estimation of standard errors, only those cases that change either on the dependent or on the independent variable in one of the observed years are used in the model.

As a robustness check, I also estimated a rare-events logistic regression (Firth, 1993; Heinze & Schemper, 2002) because the distribution of the dependent variable is strongly misbalanced (see Table 4). For both models, results point towards the same conclusion (results available upon request).

4. Results

A look at the distribution of educational levels in the working sample indicates that most respondents have an academic degree (72 per cent) and are therefore highly skilled (Table 4), potentially because refugees who make it to Europe are positively selected (Chiswick, 1999; Lange & Pfeiffer, 2018) and vocational training is less common in the countries of origin. On average, respondents, to the time of the interview, have lived in Germany for a little longer than two years (not displayed as a table). Out of the working sample, one-third had credentials for a regulated profession in Germany and 18 per cent applied for recognition. Seven per cent

were already successful in this process and received at least partial recognition. Almost 50 per cent have a secure legal title to stay in Germany and 70 per cent have already attended integration or language classes.

Table 4: Descriptive Results

	Absolute (percent)
(1) University, College and higher	1,001 (72)
(2) Vocational training	397 (28)
Applied for recognition = yes	252 (18)
Successful recognition = yes	93 (7)
Secure status = yes	680 (49)
Attending integration courses = yes	1,016 (73)
Profession regulated = yes	432 (31)

Further, we see that those who participated in integration classes received more help with their application process and are more proficient in the German language (table 5).

Table 5: Correlation between Participation in Integration Class and getting Help with Recognition and German Language Abilities (percent in parenthesis)

	No Integration Class	**Integration Class**
Help with recognition		
Got help	58 (16)	252 (26)
No help, but needed	120 (33)	321 (33)
No help needed	181 (50)	408 (42)
Cramer's V = 0.11; p = 0.00		
German language abilities		
Good	144 (38)	583 (57)
Intermediate	75 (17)	226 (22)
Poor	162 (43)	207 (20)
Cramer's V = 0.23; p = 0.00		

In the following step, I estimate multivariate models (Table 6) testing the effect of a secure status and the participation in integration/language classes on the application for degree recognition.

Table 6 displays two different multivariate models. Both models are expressed using average marginal effects (AME). The first model estimates the influence of a secure residence status and the second model the participation in integration/language classes on the application for recognition of degrees.

4.1 Humanitarian Protection Provides Incentives for Integration

Table 6: Regression Analyses on the Application for Recognition

	(1) Recognition	(2) Recognition
Secure Status	0.05***	
Ref = no	(0.01)	
Classes		0.03***
Ref = no		(0.01)
Constant	-0.07	-0.07*
	(0.06)	(0.06)
R^2 (within)	0.04	0.07
Distinct Individuals	1,171	1,303

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
Note: Both models control for each month of the panel and are weighted by using the exact matching weight. To avoid bias due to pre-treatment trends, I exclude cases that received recognition before a decision on their asylum status was made or they participated in an integration class.

For the first hypothesis, we turn to the first model in table 6. It shows a significant positive effect size with regard to the residence permit, meaning that, on average, people with a secure legal status have a 5 per cent higher likelihood of applying for recognition (AME = 0.05). Even though the effect size is rather small, this confirms the first hypothesis.

With regard to the rationale of this effect, I argue that only those who can stay longer than a year will take extra and time intensive measures in order to facilitate labour-market access. For them, the costs of the recognition procedure (around 600€ and time) are more likely to pay off in terms of, for example, a higher income and a secure job. Migrants who know that they will stay for a while are arguably more willing to invest in their economic integration, because they know that they have more time to reap the benefits of their investment. Those who are likely to be deported soon do not have the time to take intensive measures. In line with previous research, the results support the assumption that a secure legal status seems to be an incentive for taking integrational measures. In light of these findings, it is valuable to discuss the allocation of refugee protection even further and how a misallocation can impact integration trajectories.

For instance, during the peak of the inflow of refugees to Germany in 2015 and 2016, many (especially Syrian) asylum seekers were granted subsidiary protection–which is valid for one year. This was despite the fact that many of them, in theory, had the right to obtain asylum or refugee status (which is valid for three years). Besides fleeing from war, they also feared persecution. Therefore, around half of all legal actions to turn the subsidiary protection into asylum are successful (Bundesregierung der Bundesrepublik Deutschland, 2018). With regard

to this analysis, such a misallocation of asylum titles might lead to a decreasing integration effort by those refugees, thus hindering economic integration.

Furthermore, many refugees in Germany who are granted a not secure status get a prolongation of that status regularly because the situation in their home country is not improving (e.g. the crises in Iraq started in 2003). However, the uncertainty of whether they can stay in Germany for longer than one year at a time still remains. Consequently, they have only limited planning reliability and measures for long-term economic integration are presumably used rarely despite the long length of stay.

4.2 Integration Classes Provide Opportunities

For the second hypothesis, I consider the second model in Table 6. The effect size of the dummy indicating participation in integration classes and its significance strongly suggest that having visited integration and language classes is positively correlated with the application for recognition (AME = 0.03). If refugees attend integration/language classes, they are 3 per cent more likely to apply for the recognition of their degree. However, the effect is rather small. Nevertheless, Hypothesis 2 is also confirmed. If, again, we consider the descriptive results in Table 5, we have a strong indication that this is due to the network refugees build and the local language abilities acquired, because integration-class participants report that they need less help for the recognition process and have a higher German proficiency. In sum, the data suggests that integration and language classes foster economic integration.

However, empirically, the access to such classes is dependent on where the respondent lives. In some urban areas, classes are often overcrowded, while classes are infrequently offered in many rural areas because there are few potential participants. Although people with humanitarian protection should theoretically have comprehensive access, in practice, they often do not. With the results of my analysis in mind, it could be valuable to ensure that all people with humanitarian protection have actual access.[4] However, providing more classes could increase costs for a short period. Nevertheless, the data indicates that such classes can help with economic integration and, therefore, in the long run, might relieve social-welfare systems. Further, other research shows that participation in integration classes significantly improves the language proficiency of migrants, which in turn facilitates integration in general (Beenstock, Chiswick, & Repetto, 2001; Gonzalez, 2000; Hayfron, 2001; Tubergen, 2010;

[4] For example, in Germany only refugees and people with subsidiary protection are entitled to participate in integration classes. Furthermore, asylum seekers only get access when they come from a country with a good prospect to stay ("gute Bleibeperspektive") – meaning that more than fifty percent of asylum seekers from the same country of origin got a refugee status in the past.

Tubergen & Wierenga, 2011). In addition, the investment in language proficiency of migrants can be seen as a deed for the whole society because it facilitates inter-cultural communication.

5. Conclusion and Outlook

This article starts with the assumption that the institutional framework fundamentally defines the decision to apply for recognition. I further argue that the application for degree recognition is an additional proxy for economic integration because it describes the effort to transfer human capital from a foreign to a local labour market. By employing a linear-probability model combined with coarsened exact matching, I show that secure legal status and attending integration/language courses increase the likelihood of applying for the recognition of credentials. More abstract, the analyses indicate that providing migrants with planning reliability is a strong incentive for them to invest in their economic integration. These effects hold when controlling for a set of individual factors like motivational aspects, country of origin or language abilities. Thus, institutions need to be acknowledged when analysing economic integration.

Referring to the signalling theory again (Spence, 1973), I hypothesize that a recognized degree in turn facilitates labour-market access. Therefore, in future research, when more refugees have successfully applied for recognition, it could be valuable to analyse labour-market effects such as employment directly.

5.1 Outlook: Labour-market Effects of Recognized Credentials

To date, there have been only a few studies that deal with labour-market effects of credential recognition. Brücker et al. (2014), observing labour-migrant cohorts after the Second World War in Germany, found positive effects, whereas Kogan, regarding migrants from the former Soviet Union in Germany, provides mixed evidence. Her analyses indicate that credential recognition only pays off for highly skilled individuals (Kogan, 2012).

To the time of the interviews used in my own analyses in 2016, 2017 and early 2018, most of the refugees were still in the process of getting acquainted. Although some have already applied for recognition, most of them are still awaiting a decision (see again Table 4). Thus, labour-market effects will only be observable within the next few years.

Theoretically, I assume that the recognition of degrees rather improves long-term than short-term labour market access because the administrative process of recognition takes some time. Moreover, current research suggests that refugees often find their first employment through social networks (Eisnecker & Schacht, 2016): when asylum seekers arrive in Germany, the official support by the German government for finding a job or getting special training starts

only after the asylum claim is processed. Due to the fact that this process, especially during the peak migration flows in autumn 2015, could take longer than three months (after that period, asylum seekers are allowed to work), asylum seekers were pushed to find individual solutions. Eisnecker and Schacht provide indication that a private support network emerged, helping asylum seekers to find jobs within the migrant community (Eisnecker & Schacht, 2016). Within this community, the official recognition of degrees arguably is less important.

I further hypothesize that not having a recognized degree might be appealing to gatekeepers, in the sense that it places refugees in a weak position when negotiating over salary and labour conditions. Therefore, in some cases, gatekeepers might prefer those who do not have a formal degree as a way of limiting labour costs. Additionally, many formally 'unskilled' workers might still find a job in the low wage sector and, therefore, there is a market that sees 'untrained'–rather uncertified–refugees as an asset. Earlier research on this matter indicates that, in some cases, it can be rational to invest in quick labour-market access instead of taking the detour of credential recognition as a first step. Kogan shows, for instance, that a recognition of degrees is less useful for migrants with low qualifications (Kogan, 2012).

Thus, recognition of certificates might not play a key role for initial labour-market access. However, it would be worthwhile analysing whether long-term labour-market integration benefits from recognized degrees. In this case, I hypothesize that, if a degree is recognized, thus providing a signal of economic utility, the probability of a job mismatch in the future will decline. Having a recognized degree should help during job negotiations, because proof exists that the degree is valuable in the German context. Assuming that this helps to facilitate the avoidance of education mismatch, it provides a higher income (Nordin, Persson, & Roth, 2010). Therefore, although labour-market access itself might not be supported, recognized degrees should help in finding a job that matches one's profession. Unfortunately, and due to the fact that the current cohort of refugees just recently arrived in Germany, long-term labour-market access with respect to degree recognition cannot be studied yet.

5.2 Limitations

A clear advantage of this study compared to those relying on the same dataset is that I carry out panel fixed-effects regression on more than two years. Combined with the matching, I come close to a set-up that allows causal interpretation. However, matching can be carried out with observable variables, only. Therefore, unobserved time-variant confounders might be an issue, tackling causality. Unfortunately, this problem regularly emerges when relying on survey data, but, similar conclusions have been presented in other studies on naturalization that

relied on experimental data (Hainmueller et al., 2017). Therefore, bias due to unobserved heterogeneity should not be a big issue.

Because the treatment in the model is not random, reverse causality might be a problem as well. Using this data and design, reverse causality cannot be ruled out in total. However, by omitting cases from the analyses where the application for recognition was obeserved prior to getting a safe asylum status (N = 116 (8%)) or participating in classes (N = 92 (7%)) this issue does not substantially tackle my findings. Additionally, referring to previous research and theoretical assumptions, I am even more confident that this issue is rather small. For instance, it is implausible that the application for recognition of degrees leads to the desire to take legal actions against an insecure residence status. Both steps are institutionally not intertwined. Furthermore, it is unlikely that official integration classes teach about legal actions against unsecure residence permits. Because most classes are organized by the BAMF, it is implausible that they advocate against their own regulations.

By choosing the recognition of degrees as the dependent variable, I am able to tackle a widely discussed problem and phenomenon in migration research: transferability of human capital. As I did not choose the result of this procedure, but rather the decision for recognition as the dependent variable, I am able to analyse the decision of individual actors within an institutional framework. The results indicate that an institutional framework that provides opportunities leads to an increasing effort of economic integration. However, as some refugees do not hold formal degrees, this analysis cannot make statements about trained but uncertified refugees. Therefore, further research needs to examine the situation of those migrants who are not able to provide proof of education and training, investigating how institutions shape their economic integration.

The integration of recent refugees has not yet been thoroughly analysed within the research literature. This study provides first insights into the importance of the institutional framework and points out that a secure residence status, and to a lesser extent, the participation in integration/language classes is crucial for understanding integration trajectories.

References

Ager, A., & Strang, A. (2008). Understanding Integration: A Conceptual Framework. *Journal of Refugee Studies, 21*(2), 166-191.

Australian Government - Department of Education. (2019). Qualification assessment. Retrieved from https://internationaleducation.gov.au/services-and-resources/services-for-individuals/qualifications-assessments/pages/qualification-assessment.aspx, last visited: 02.04.2020

Bakker, L., Dagevos, J., & Engbersen, G. (2017). Explaining the refugee gap: a longitudinal study on labour market participation of refugees in the netherlands. *Journal of Ethnic and Migration Studies, 43*(11), 1775-1791. doi:10.1080/1369183X.2016.1251835

Bakker, L., Degevos, J., & Engbersen, G. (2013). The Importance of Resources and Security in the Socio-Economic Integration of Refugees. A Study on the Impact of Length of Stay in Asylum Accommodation and Residence Status on Socio-Economic Integration for the Four Largest Refugee Groups in the Netherlands. *Journal of International Migration and Integration*, 1-18. doi:DOI 10.1007/s12134-013-0296-2

BAMF. (2016a). Forms of Protection. Retrieved from http://www.bamf.de/EN/Fluechtlingsschutz/AblaufAsylv/Schutzformen/schutzformen-node.html, last visited: 01.10.2018

BAMF. (2016b). Integration courses for asylum applicants. Retrieved from http://www.bamf.de/EN/Willkommen/DeutschLernen/IntegrationskurseAsylbewerber/integrationskurseasylbewerber-node.html, last visited: 01.10.2018

Becker, G. S. (1993). *Human Capital: A Theoretical and Empirical Analysis, With Special Reference to Education*. Chicago: University of Chicago Press.

Beenstock, M., Chiswick, B. R., & Repetto, G. L. (2001). The effect of linguistic distance and country of origin on immigrant language skills: application to Israel. *International Migration, 39*(3), 33-60. doi:doi:10.1111/1468-2435.00155

Bellamy, R. (2008). *Citizenship: A Very Short Introduction*. New York: Oxford Universtiy Press.

Berry, J. W. (1997). Immigration, Acculturation, and Adaption. *Applied Psychology: an International Review, 46*(1), 5-68.

Bevelander, P. (2011). The Employment Integration of Resettled Refugees, Asylum Claimants, and Family Reunion Migrants in Sweden. *Refugee Survey Quarterly, 30*(1), 22-43.

Blossfeld, H.-P., & Mayer, K. U. (1988). Labor Market Segmentation in the Federal Republic of Germany: an Empirical Study of Segmentation Theories from a Life Course Perspective. *Europea Sociological Review, 4*(2), 123-140. doi:10.1093/oxfordjournals.esr.a036472

BMBF. (2017). *Bericht zum Anerkennungsgesetz 2017*. Bundesministerium für Bilduing und Forschung.

BMBF. (2019). Was kostet das Anerkennungsverfahren? Retrieved from https://www.anerkennung-in-deutschland.de/html/de/faq.php, last visited: 02.04.2020

Borjas, G. J. (1991). Immigration Policy, National Origin, and Immigrant Skills: A Comparison of Canada and the Unidted States. In D. Card & R. B. Freeman (Eds.), *Small Differences That Matter: Labor Markets and Income Maintenance in Canada and the United States* (pp. 21-44). Chicago: University of Chicago Press.

Brücker, H., Liebau, E., Romiti, A., & Vallizadeh, E. (2014). Anerkannte Abschlüsse und Deutschkenntnisse lohnen sich. *IAB-Kurzbericht, 21*(3), 21-28.

Brücker, H., Rother, N., & Schupp, J. (2016). *IAB-BAMF-SOEP-Befragung von Geflüchteten: Überblick und erste Ergebnisse*.

Brücker, H., Rother, N., & Schupp, J. (2017). IAB-BAMF-SOEP-Befragung von Geflüchteten 2016: Studiendesign, Feldergebnisse sowie Analysen zu schulischer wie beruflicher Qualifikation, Sprachkenntnissen sowie kognitiven Potenzialen. *Politikberatung kompakt, 123*.

Brussig, M., Dittmer, V., & Knuth, M. (2009). Verschenkte Potenziale: Fehlende Anerkennung von Qualifikationsabschlüssen erschwert die Erwerbsintegration von ALG II-Bezieher/innen mit Migrationshintergrund. *IAQ Report, 2009*(8).

Bundesministerium Europa Integration und Äußeres. (2019). Overview: Procedure for formal recognition. Retrieved from https://www.berufsanerkennung.at/en/professional-recognition/overview/, last visited: 02.04.2020

Bundesregierung der Bundesrepublik Deutschland. (2018). *Bundestagsdrucksache 19/385*.

Cheung, S. Y., & Phillimore, J. (2014). Refugees, Social Capital, and Labour Market Integration in the UK. *Sociology, 48*(3), 518-536.

Chin, A., & Cortes, K. E. (2015). The Refugee/Asylum Seeker. In B. R. Chiswick & O. W. Miller (Eds.), *Handbook of the Economics of International Migration* (Vol. 1A, pp. 585-658). Oxford: Elsevier.

Chiswick, B. R. (1999). Are immigrants favorably self-selected? *American Economic Review, 89*(2), 181-185.

Chiswick, B. R., Lee, Y. L., & Miller, P. W. (2006). Immigrants' Language Skills and Visa Category. *International Migration Review, 40*(2), 419-450.

Chiswick, B. R., & Miller, P. W. (1992). Language in the immigrant labour market. In B. R. Chiswick (Ed.), *Immigration, Language and Ethnicity: Canada and the United States* (pp. 229-296). Washington, DC: American Enterprise Institute.

Chiswick, B. R., & Miller, P. W. (2009). The international transferability of immigrants' human capital. *Economics of Education Review, 28*(2), 162-169.

Delanty, G., Jones, P., & Wodak, R. (2008). Introduction: Migration, Discrimination and Belonging in Europe. In G. Delanty, P. Jones, & R. Wodak (Eds.), *Identity, Belonging and Migration* (pp. 1-20). Liverpool: Liverpool University Press.

Dumont, J.-C., Liebig, T., Peschner, J., Tanay, F., & Xenogiani, T. (2016). *How are refugees faring on the labour market in Europe?* Brussels: European Commission & OECD

DuMouchel, W. H., & Duncan, G. J. (1983). Using Sample Survey Weights in Multiple Regression Analyses of Stratified Samples. *Journal of the American Statistical Association, 78*(383), 535-543. doi:10.2307/2288115

Eisnecker, P. (2019). Non-Migrants' Interethnic Relationships with Migrants: The Role of the Residential Area, the Workplace, and Attitudes toward Migrants from a Longitudinal Perspective. *Journal of Ethnic and Migration Studies, 45*(5), 804-824. doi:10.1080/1369183X.2017.1394180

Eisnecker, P., & Schacht, D. (2016). Die Hälfte der Geflüchteten in Deutschland fand ihre eerste Stelle über soziale Kontakte. *DIW Wochenbericht, 2016*(35), 757-764.

Esser, H. (2006). *Sprache und Integration - Die sozialen Bedingungen und Folgen des Spracherwerbs von Migraten*. Frankfurt/New York: Campus Verlag.

Favell, A. (1998). *Philosophy of Integration. Immigration and the Idea of Citizenship in France and Britain*. Basingstoke: Macmillan.

Firth, D. (1993). Bias reduction of maximum likelihood estimates. *Biometrika, 80*(1), 27-38. doi:10.1093/biomet/80.1.27

Friedberg, R. M. (2000). You can't take it with you? Immigrant assimilation and the protability of human capital. *Journal of Labor Economics, 18*(2), 221-251.

Göbel, J., Grabka, M. M., Liebig, S., Kroh, M., Richter, D., Schröder, C., & Schupp, J. (2019). The German Socio-Economic Panel Study (SOEP). *Jahrbücher für Nationalökonomie und Statistik*(238).

Gonzalez, A. (2000). The Acquisition and Labor Market Value of Four English Skills: New Evidence from NALS. *Contemporary Economicy Policy, 18*(3), 259-269. doi:10.1111/j.1465-7287.2000.tb00023.x

Gostomski, C. B. v., & Pupeter, M. (2008). Zufallsbefragung von Ausländern auf Basis des Ausländerzentralregisters. *Methoden, Daten, Analysen, 2*(2), 149-177.

Government of Canada. (2019). Funding: Foreign Credential Recognition Program - Overview. Retrieved from https://www.canada.ca/en/employment-social-

development/services/funding/foreign-credential-recognition.html, last visited: 02.04.2020
Hainmueller, J., & Hangartner, D. (2013). Who gets a Swiss passport? A Natural Experiment in Immigrant Discrimination. *American Political Science Review, 107*(1), 159-187. doi:doi:10.1017/S0003055412000494
Hainmueller, J., Hangartner, D., & Pietrantuono, G. (2015). Naturalization Fosters the long-Term Political Integration of Immigrants. *Proceedings of the National Academy of Science of the United States of America, 112*(41), 12651-12656. doi:DOI: 10.1073/pnas.1418794112
Hainmueller, J., Hangartner, D., & Pietrantuono, G. (2017). Catalyst or Crown: Does Naturalization Promote the Long-Term Social Integration of Immigrants? *American Political Science Review, 111*(2), 256-276. doi:10.1017/S0003055416000745
Hale, S. (2000). The Reception and Resettlement of Vietnamese Refugees in Britain. In V. Robinson (Ed.), *The International Refugee Crisis* (pp. 280-290). Basingstoke: Macmillan.
Hall, P. A. (2010). Historical Insitutionalism in Rationalist and Sociological Perspective. In J. Mahoney & K. Thelen (Eds.), *Exaplaining Institutional Change: Ambiguity, Agency, and Power* (pp. 204-224). Cambridge: Cambridge University Press.
Hayfron, J. E. (2001). Language training, language proficiency and earnings of immigrants in Norway. *Applied Economics, 33*(15), 1971-1979. doi:10.1080/00036840010018630
Heinze, G., & Schemper, M. (2002). A solution to the problem of separation in logistic regression. *Statistics in Medicine, 21*(16), 2409-2419. doi:10.1002/sim.1047
Hoehne, J., & Michalowski, I. (2016). Long-Term Effects of Language Course Timing on Language Acquisition and Social Contacts: Turkish and Moroccan Immigrants in Western Europe. *International Migration Review, 50*(1), 133-162. doi:10.1111/imre.12130
Imbens, G. W., & Wooldridge, J. M. (2009). Recent developments in the econometrics of program evaluation. *Journal of Economic Literature, 47*(1), 5-86. doi:http://dx.doi.org/10.1257/jel.47.1.5
Imdorf, C. (2010). Wie Ausbildungsbetriebe soziale Ungleichheit reproduzieren: Der Ausschluss von Migrantenjugendlichen bei der Lehrlingsselektion. In H.-H. Krüger (Ed.), *Bildungsungleichheit Revisited* (pp. 259-274). Wiesbaden: VS Verlag für Sozialwissenschaften.
Jacobsen, J., Kroh, M., Kühne, S., Scheible, J. A., Siegers, R., & Siegert, M. (2019). Supplementary of the IAB-BAMF-SOEP Survey of Refugees in Germany (M5) 2017. *SOEP Survey Papers Series C - Data Documentation*(605).
Jacobsen, J., Kroh, M., Legewie, N., & Salikutluk, Z. (2018). Arbeitsmarktpotenzial, Qualifikationsniveau, berufliche Perspektiven Geflüchteter. In H. Floeting, G. Kirchhoff, H. Scheller, & J. Schneider (Eds.), *Zuwanderung und Integration von Geflüchteten in Kommunen: Der lange Weg vom Ankommen zum Bleiben* (pp. 99-105). Berlin: DifU.
Jones, P., & Krzyzanowski, M. (2008). Identity, Belonging and Migration: Beyond Constructing "Others". In G. Delanty, R. Wodak, & P. Jones (Eds.), *Identity, Belonging and Migration* (pp. 38-53). Liverpool: Liverpool University Press.
Knight, J. (1992). *Institutions and Social Conflict*. New York: Cambridge University Press.
Kogan, I. (2012). Potenziale nutzen! Determinanten und Konsequenzen der Anerkennung von Bildungsabschlüssen bei Zuwanderern aus der ehemaligen Sowjetunion in Deutschland. *Kölner Zeitschrift für Soziologie und Sozialpsychologie, 64*(1), 67-89. doi:10.1007/s11577-012-0157-6
Koopmans, R. (2009). Trade-Offs between Equality andDifference: Immigrant Integration, Multiculturalism and the WelfareState in Cross-National Perspective. *Journal of Ethnic and Migration Studies, 36*(1), 1-26. doi:10.1080/13691830903250881
Krahn, H., Derwing, T., Mulder, M., & Wilkinson, L. (2000). Educated and Underemployed: Refugee Integration into the Canadian Labour Market. *Journal of International Migration and Integration, 1*(1), 59-84.

Kroh, M., Kühne, S., Jacobsen, J., Siegert, M., & Siegers, R. (2017). Sampling, Nonresponse, and Integrated Weighting of the 2016 IAB-BAMF-SOEP Survey of Refugees (M3/M4). *SOEP Survey Papers Series C - Data Documentation*(477).

Lange, M., & Pfeiffer, F. (2018). The Human Capital Selection of Young Males Seeking Asylum in Germany. *ZEW Discussion Paper, 18*(40), 1-22. doi:http://hdl.handle.net/10419/183151

Nordin, M., Persson, I., & Roth, D.-O. (2010). Education-occupation mismatch: Is there an income penalty? *Economics of Education Review, 29*(6), 1047-1059. doi:10.1016/j.econedurev.2010.05.005

Ohliger, R., & Brands, F. (2016). *Chancen Erkennen - Perspektiven schaffen - Integration ermöglichen*. Robert-Bosch Stifung

Phillimore, J., & Goodson, L. (2006). Problem or Opportunity? Asylum Seekers, Refugees, Employment and Social Exclusion in Deprived Urban Areas. *Urban Studies, 43*(10), 1715-1736.

Pietkaza-Nykaza, E. (2014). "I want to do anything which is Decent and Relates to My Profession": Refugee Doctors' nd Teachers' Strategies of Re-Entering Their Profession in the UK. *Journal of Refugee Studies, 28*(4), 523-543.

Pietrantuono, G. (2016). *The Value of Citizenship. Experimental and Quasi-experimental Evidence from Germany and Switzerland.* (PhD Dissertation). Universtiät Mannheim, Mannheim.

Portes, A., & Rumbaut, R. G. (2001). *Legacies: The story of the Immigrant Second Generation*. New York: Russal Sage Foundation.

Pries, L. (2016). *Migration und Ankommen*. Frankfurt/New York: Campus Verlag.

Scheinman, R. S. (1983). Refugees: Goodbye to the good old days. *The Annals*(467), 78-88.

Schmitz, N. (2018). *Auswertung der amtlichen Statistik zum Anerkennungsgesetzdes Bundes für 2017*. Bundesinstitut für Berufsbildung

Schuller, K. (2011). Der Einfluss des Integrationskurses auf die Integratino russisch- und türkischstämmiger Integrationskursteilnehmerinnen. *Working Paper der Forschungsgruppe des Bundesamtes*(37).

Seibert, H., & Solga, H. (2005). Gleiche Chancen dank einer abgeschlossenen Ausbildung? Zum Signalwert von Ausbildungsabschlüssen bei ausländischen und deutschen jungen Erwachsenen. *Zeitschrift für Soziologie, 34*(5), 364-382.

Sicherman, N. (1991). "Overdeducation" in the Labor Market. *Journal of Labor Economics, 9*(2), 101-122.

Spence, M. (1973). Job Market Signaling. *Quarterly Journal of Economics, 87*(3), 355-374.

Spence, M. (1978). Job Market Signaling. In P. Diamond & M. Rothschild (Eds.), *Uncertainty in Economics* (pp. 283-306). New York: Elsevier.

Swedish Council for Higher Education. (2019). Recognition of foreign qualifications. Retrieved from https://www.uhr.se/en/start/recognition-of-foreign-qualifications/, last visited: 02.04.2020

Tubergen, F. v. (2010). Determinants of Second Language Proficiency among Refugees in the Netherlands. *Social Forces, 89*(2), 515-534. doi:10.1353/sof.2010.0092

Tubergen, F. v., & Wierenga, M. (2011). The Language Acquisition of Male Immigrants in a Multilingual Destination: Turks and Moroccans in Belgium. *Journal of Ethnic and Migration Studies, 37*(7), 1039-1057. doi:10.1080/1369183X.2011.572476

Convention and Protocol Relating to the Status of Refugees, (1966).

Verbruggen, M., Emmerik, H. v., Gils, A. V., Meng, C., & Grip, A. d. (2015). Does Early-Career Underemployment Impact Future Career Success? A Path Dependency Perspective. *Journal of Vocational Behavior, 90*, 101-110. doi:10.1016/j.jvb.2015.08.002

Weiermair, K. (1976). The Economic Effects of Language Training to Immigrations: A case study. *International Migration Review, 10*(2), 205-219. doi:10.1177/019791837601000204

Appendix

Table A1: Defining secure and not secure residence permits

Secure	Not secure
Residence Permit (Article25 Section1 & 2 Residence Act) / Aufenthaltserlaubnis (§25 Abs. 1 & 2 Aufenthaltsgesetz)	**Temporary Residence Permit** (Article55 Asylum Act) / Aufenthaltsgestattung (§55 Asylgesetz)
Permanent residency (Article26 Section3 Residence Act) Niederlassungserlaubnis (§25 Abs. 3 Aufenthaltsgesetz)	**Suspension of deportation** (Article 60a Residence Act) / Duldung (§60a Aufenthaltsgesetz)
	Residence Permit (Article 22 or 23 Residence Act) Aufenthaltserlaubnis (§22 / §23 Aufenthaltsgesetz)
	Residence Permit (Article 25 Section 3, 4 or 5 Residence Act) Aufenthaltserlaubnis (§25 Abs. 3 / 4 / 5 Aufenthaltsgesetz)
	Other

Table A2: Possible Confounding Factors in Multivariate Regression

Motivation coming to Germany. 1) Because of labor market or educational system 2) other
Years of education 1) 1-9 years 2) 10-13 years 3) 13+ years
German proficiency 1) good 2) medium 3) bad
Current age in terciles
Mother tongue available on anerkennung-in-deutschland.de? 1) Yes 2) no
German proficiency 1) good 2) medium 3) bad
Country of Origin with good "prospect of remaining in the country"?[1] 1) yes 2) no
Profession is formally regulated in Germany 1) yes 2) no

[1] Note: The Federal Office for Migration and Refugees classifies asylum seekers with regard to their prospect of obtaining asylum or refugee protection. Currently, those chances are high for respondents Syria, Eritrea, Iraq, Iran, and Somalia. This classification takes into account how many people from a respective country are provided with asylum. If this rate is higher than 50%, the chances are officially declared as "good". This classification serves as a reference of whether certain people get early access to integration classes for the time of their asylum application.

Table A3: Regression on Secure Status and Participation in Integration/Language Class

	(1) Secure Status	(2) Class
Motivation to stay in Germany = yes	1.566***	1.324
	(0.20)	(0.19)
9 years schooling	1	1
	(.)	(.)
13 years schooling	1.495	1.682*
	(0.31)	(0.36)
	1.495	1.682*
	(0.31)	(0.36)
More than 13 years of schooling	1.131	1.415
	(0.27)	(0.36)
CoO with good prospect of staying = yes	1.231	2.279***
	(0.20)	(0.40)
Good German	1	1
	(.)	(.)
Intermediate German	1.037	0.693*
	(0.17)	(0.13)
Poor German	1.631**	0.292***
	(0.25)	(0.05)
3 quantiles of immiage[1]=1	1	1
	(.)	(.)
3 quantiles of immiage=2	1.372	0.756
	(0.51)	(0.30)
3 quantiles of immiage=3	1.277	2.392
	(0.66)	(1.34)
3 quantiles of age=1	1	1
	(.)	(.)
3 quantiles of age=2	0.471*	1.729
	(0.18)	(0.70)
3 quantiles of age=3	0.489	0.887
	(0.25)	(0.49)

Mother tongue provided on web page? = yes	0.713* (0.10)	0.733 (0.12)
Profession regulated = Yes	1 (.)	1 (.)
No	0.924 (0.14)	1.093 (0.20)
Profession not clear	1.083 (0.18)	0.938 (0.18)
Pseudo R^2	0.029	0.089
Observations	1149	1149

Exponentiated coefficients; Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

[1] Age at time of immigration

8 Conclusion

This thesis took the IAB-BAMF-SOEP Survey of Refugees as an example, explored challenges during its implementation, and applied lessons learned to the measurement of refugee integration. Four challenges were identified: Sampling, provision of translated questionnaires, cross-cultural comparability of latent constructs, and applying target population specific assumption on integration. Those four challenges were described and analyzed in four different articles:

1) In the first article, my colleagues and I identified that for sampling refugees in times of high immigration a sequential sampling strategy is a good fit for time-lagged sampling frames.

2) In a second article, I show that a language mismatch increases item nonresponse and that audio recordings cannot diminish this effect.

3) The third article indicates that political culture and language limit the comparability of conceptions of democratic values.

4) In the fourth article, I argue that the institutional framework for refugees is a crucial aspect to consider when studying economic integration. To show this, I analyze the effect of secure residence permits and the attendance in integration classes on the decision to apply for recognition of vocational certificates.

Table 1 summarizes the empirical chapters.

Table 1: Summary of Empirical Chapters

	Chapter 5	Chapter 6	Chapter 7
Research Question	How does a language mismatch affect item nonresponse?	Are conceptions of democracy comparable cross culturally?	What factors drive an investment for labor market access of refugees?
Theory	Respondent Burden (Krosnick, 1991), Social Desirability (Tourangeau & Yan 2007)	Cross-Cultural Measurement Invariance of Democratic Attitudes (Ariely and Davidov 2011), The Structure of Social Action (Parsons, 1965), Clash of Civilization (Huntington, 1993; Inglehart & Norris, 2003)	Indicators of Integration (Ager and Strang 2008)
Hypotheses	1) A mismatch between survey language and item nonresponse increases item nonresponse 2) Applying audio files diminishes item nonresponse	1) Conceptions of democracy are not comparable across countries of origin 2) Conceptions of democracy are not comparable across languages	1) Refugees with a safe resident permit have a higher chance of applying for a recognition of their vocational degree 2) Refugees who participated in integration classes have a higher chance of applying for a recognition of their vocational degree.
Methods	Zero Inflated Poisson Regression and Multi-Level Linear Probability Model	Multiple Group Confirmatory Factor Analyses	Linear Regression Analyses and Coarsened Exact Matching
Key Findings	Not providing questionnaires in a respondent's mother tongue increases the chances of item nonresponse. The use of audio files in surveys does not decrease item nonresponse.	The understanding of democracy is likely not comparable between refugees and Germans. The understanding of democracy is likely not comparable between refugees from different countries. The understanding of democracy is likely not comparable across languages.	A secure residence permit leads to a higher chance that refugees will apply for a recognition of their credentials. The participation in language/integration classes leads to a higher chance that refugees will apply for recognition of their credentials.

As with many survey methodological and theoretical considerations, it is usually impossible to implement all recommendations comprehensively. Therefore, even though this thesis suggests many solutions for the indicated challenges, they must be viewed in light of their applicability. One factor that hampers implementation are survey costs. I recommend in Chapter 5, for example, that in the future, survey institutes should always provide questionnaires in all respondents' mother tongues. In the case of the IAB-BAMF-SOEP Survey of Refugees, this would have implied translation into over 20 different languages, some of which would have been used only scarcely. Therefore, in order to balance costs and benefits, it might be rational to rely on the most important ones in order to diminish the bias. However, consequences such as item nonresponse should then be implemented in the analyses and the communication of results.

A similar obstacle occurs regarding comparability: For instance, if a survey on refugees is designed to comprehensively reflect the specificities of refugees, this might impede its integration into a larger general population survey, as instruments differ. For example, as shown in Chapter 6 some latent constructs are likely not to be comparable across country of origin or language. However, to date, no items exist that would overcome such problems. Some solutions have been described such as culturally sensitive development of manifest variables (i.e., varying items across groups), which nevertheless reflect the same latent construct statistically (for an example see Boehnke, 2018). However, these efforts are still in their infancy. Therefore, it might be valuable to implement known but problematic variables nevertheless. Having said this, and with Chapter 6 in mind, researchers should then be careful when interpreting these data and always keep in mind that to date there is no universal indicator of democratic conceptions, but rather normative assumptions about what democracy should look like.

In light of this trade-off between statistical accuracy and applicability, this thesis provides some broader implications for future research:

1) Time lagged sampling frames do not pose a challenge to random samples.
2) Item nonresponse in cross-cultural survey reflect language issues.
3) Audio files do not decrease the burden on respondents.
4) Latent constructs need to be tested for measurement invariance.
5) Institutions are a crucial factor in integration processes.

Moving away from a purely academic conclusion, this thesis also entails some contributions to a broader public and policy-related debate on refugee integration.

Political and Societal Implications

The first two papers regarding sampling and language mismatch could also be viewed in light of how data and social reporting is presented in the public sphere. Both papers show difficulties that emerge during the survey process. Throughout the sampling and field phase of the survey process, these problems can be mitigated. Thus, when using delicate data for policy reporting or for journalism, as has been done extensively with the IAB-BAMF-SOEP Survey of Refugees, scholars need to provide information about or correct for possible issues that might hamper the statistical accuracy of the survey. Otherwise, policy makers might rely on data that do not actually provide information about the target population but only about a selective group, and this could easily lead to faulty policy decisions.

Moreover, handling survey data inaccurately entails the risk of shaping societal perceptions on the research subject based on statistical artefacts. For example, employment rates (especially those of migrants) are a prominent and frequently discussed topic in German society, and increasing unemployment rates of refugees could have major impact on the discourse on refugee integration. Therefore, it is important that such observations are based in thorough analyses that take survey methodological aspects into account.

Examining these considerations in light of the debate on "fake news" suggests that researchers could be more transparent about the mechanisms of data production[1]. If researchers emphasize that survey results are usually based on probabilistic methods, the broader public sphere might gain a better understanding of the fact that even though some predictions do not materialize, this does not call into question survey research in general. Additionally, scholars who provide policy advice should inform policy makers that point estimates cannot be taken at face value and do not necessarily reflect the true value of the target population, but only of the sample. By using and explaining, for

[1] At this point, it is crucial to mention that I do not support the claim that academia in Germany is at risk of producing fake news. With this sentence I refer to a discourse, which questions the reliability of survey results. The debate on misleading opinion polls prior to the 2016 US election are a prominent example in that respect.

instance, confidence intervals in the public presentation of results, researchers would counteract a quasi-deterministic understanding of survey data. This in turn would promote a more accurate perception of survey results, which then might lead to more trust in social science research in general.

Besides methodological insights, this thesis also provides some conclusions from a more practical policy perspective, especially the articles presented in Chapters 6 and 7.

With the arrival of refugees since 2015, there has been increased debate over shared value conceptions, with some even raising the question whether individuals who do not share the fundamental values of German society should be granted the right to immigration (see e.g. the essay by the German minister for domestic affairs at that time, de Maizière, 2017). The results discussed here, by showing that the comparability of democratic conceptions is limited, call into question whether one can problematize the "societal" fit of refugees with core values of German society. With reference to previous research indicating that value conceptions vary within societies (Schwartz & Sagie, 2000), this thesis provides additional indications that an empirical test for differences in value conceptions cannot be carried out cross-culturally among the refugee population to date. Therefore, there are strong empirical indications that the concept of universal values is at least contested and that cross-cultural comparisons are prone to measurement error. Therefore, in the future and until adequate items exist, it should be made clear that the definition of societal values is a normative and political endeavor and not the result of empirical and evidence-based analysis. Referring to the aforementioned debate on value consensus between refugees and the host society, the claim that refugees might not "fit" cannot currently be backed up empirically. Therefore, such claims should be treated with caution.

Moreover, the article on institutional aspects of credential recognition supports the hypothesis that a prospect of remaining in the country longer increases integration efforts of refugees. Similar evidence has already been provided before with regard to other migrants (Hainmueller, Hangartner & Pietrantuono, 2015). From a political stance, it might therefore be rational to provide such safety for refugees comprehensively: as my results show that this form of safety leads to an investment in labor market access, extending the residence permits could subsequently relieve the social welfare system because refugees should be more likely to find employment.

However, this finding can also be viewed from a different angle. Because the results presented here rely on a target population that already underwent some sort of formal training in their home country, from a policy perspective one has to ask how refugees without proper certification will fare on the labor market in the future. Even more so those refugees who have no certification of vocational training, and additionally do not have a safe residence permit. Considering the results presented in Chapter 7, which dealt with skilled refugees, one can only hypothesize how much worse the prospects would be for those who have no proof of formal training. Two possible solutions for unskilled refugees come to mind. First, providing such refugees with support in the form of welfare benefits to avoid an increase in absolute poverty, or to lower the barriers of labor market access for those who have demonstrable practical skills but no official certification. The second option seems more promising. This is for two reasons: First, welfare benefits debit the public budget. Additionally, considering some current public sentiments towards refugees (e.g. Jacobsen et al., 2017), it is questionable whether such an increase in welfare expenditure would be politically enforceable. Facilitating labor market access for refugees without vocational certification would, however, probably increase employment rates among refugees and subsequently relieve the welfare state. Additionally, having in mind the debate on the lack of specialists in Germany (Fuchs et al., 2019), especially in manual work, the current inflow of refugees could also be an opportunity to fill this gap. As most refugees are likely to stay in Germany, as the war in Syria is not coming to an end and the situation in Afghanistan, Iraq, Eritrea, and Somalia remains precarious, I would assume that such an investment by German entrepreneurs would pay off in the future.

In sum, this thesis provides some valuable insights, not only for academic research but also for the public and policy makers. Based on the results presented in my thesis, I argue that survey producers should be more transparent about the generation of survey data, including its flaws. Moreover, I suggest that a debate on value consensus should be marked as normative and political, and I additionally provide indications that the provision of safe residence permits increases refugee's investment in future labor market access.

8.1 Limitations

Despite important contributions to the academic and the political sphere, some limitations of this dissertation should be noted. From a general viewpoint, I need to mention that I was unable to rely on experimental data in this thesis. Thus, I did not use observations in a controlled setting. This implies that estimated treatment effects are not based on randomization. However, due to the richness of variables in the IAB-BAMF-SOEP Survey of Refugees, I was capable to account for this by means of control variables and statistical matching to some extent.

From a more detailed perspective, some further limitations are worth mentioning (see also the limitations sections in each acrticle). First, the article on language mismatch only addresses item nonresponse. Although the choice of item nonresponse as the dependent variable was carefully considered, as it has been neglected in a multi-cultural setting to date, other effects such as measurement error could be plausible effects of language mismatch as well, a perspective not addressed in this dissertation. Moreover, the question of language in surveys could also have been connected to the role of the interviewers as intermediaries. Unfortunately, interviewer information was scare. However, a perspective on how language abilities of interviewers could improve survey quality would have been an interesting research topic as well. Additionally, this would also refer to the survey life cycle, which emphasizes the role of interviewers for maintaining high survey quality.

Second, Chapter 6 on measurement invariance solely focuses on values, in particular on democratic values. It is plausible that the findings can be applied to other value-related constructs; however, this thesis does not provide indications on this.

Third, Chapter 7 finds significant but rather small effects for participation in integration classes and a secure residence title on application for recognition of vocational qualifications. On the one hand, this raises the question of whether this result is due to a type-one error. On the other hand, comparing the effect size to the actual share of respondents who have applied for recognition of qualifications already puts the effect size in a different light (18% have applied, AME for secure status = 5%, for participation in class = 3%). Therefore, I recommend replicating this finding, first, with the same data but at a time point in the panel when more respondents have applied for recognition.

In sum, this thesis should be discussed in light of its limitations. However, as research is founded in theory, which is subsequently tested empirically, not all narratives and findings are flawless or without contradictions (see e.g. the Editorial Nature Human Behaviour, 2020). Additionally, empirical research is cumulative, meaning that new studies build on past studies, and studies in the future might therefore refer to my work. Thus, by encouraging the replication of my findings and by proposing advancements based on these limitations, I am confident that remaining uncertainties can be elucidated in the future and should not mitigate the general quality of this work.

8.2 Outlook

The limitations of this thesis point to some ideas for future research projects.

First, the question of language mismatch in a multi-linguistic survey can be extended to a broader set of errors. It is likely that not only item nonresponse increases, but also affected questions are understood differently, and thus measurement is inefficient or even biased. Therefore, in a next step, it would be valuable to estimate whether language mismatch introduces measurement variance or even bias.

Also, from a general, non-migration-studies-related perspective, the question of language is a crucial one, as native speakers also differ in their ability to speak their language (Perry & Gauly, 2019). Therefore, it would be valuable to test whether the use of simple language leads to decreasing item nonresponse in general population surveys.

Additionally, when dealing with item nonresponse, the question of social desirability in cross-cultural surveys could be emphasized more. Unfortunately, with the current design of the IAB-BAMF-SOEP Survey of Refugees, I could not implement a field experiment testing social desirability in a controlled setting. However, as part of the second phase of the project in which I wrote this thesis (the GeFam or *Geflüchtete Familien* project on refugee families) and which partly funds this survey, a focus on social desirability is part of the project proposal. Thus, as a follow-up to this thesis, I will focus on causes of social desirability. Two strategies to identify and overcome socially desirable response behavior seem plausible:

1) In order to tackle aforementioned limitations of this thesis, I would apply an experiment within the IAB-BAMF-SOEP Survey of Refugees (thus with random allocation of treatment and control). As has been shown previously, socially

desirable response behavior can be decreased by self-interviewing (Kühne 2018; Nederhof 1985). Therefore, the treatment group could fill out potentially sensitive questions themselves, without the interviewer looking. Thus, a main cause of social desirability (presence of a third party) would be overcome. In a second step, I would analyze whether self-administered questionnaires produce less item nonresponse. If this is the case, sensitive questions should be answered via self-interviewing.

2) A second strategy, which does not need randomization and self-interviewing, would be the cross-wise model, in which a known sensitive question is combined with a trivial one of which the distribution is known (e.g., month of birth of mother). The two questions are asked simultaneously, for instance, "Did you use hard drugs in the last week?" and "Was your mother born after June?" Subsequently, respondents only indicate whether they answered one question with yes, or both with no. Because the distribution of months of birth is known, we could identify the share of respondents who ticked yes on a sensitive question (e.g. Jann, Jerke & Krumpal, 2011) without putting the respondent in the position of actually answering the sensitive question.

Identifying socially desirable response patterns of refugees could then complement the refugee-specific analysis of item nonresponse. Moreover, applying the cross-wise model would bear the potential to include sensitive topics in the survey without switching between face-to-face and self-interviewing.

Regarding Chapter 6 and the question of measurement invariance, I suggest two consecutive research ideas:

1) Chapter 6 ends with the conclusion that the items presented are currently not capable of estimating differences in value conceptions, and proposes developing new questions by means of qualitative interviews. However, it would be interesting to test whether this assumption holds over time. It seems plausible that the reasons for measurement variance (political culture, socialization) align over time and thus the tested instruments improve. In order to do this, the same questions could be replicated in future waves or in other surveys with migrants who are already covered longer than the refugees in the IAB-BAMF-SOEP Survey of Refugees. Doing this would also contribute to the integration literature in

general, as it would test the hypothesis of classical assimilation theory which states that newcomers and locals align culturally over time until differences are no longer visible.

2) Additionally, it would be important to show the comparability of other crucial constructs implemented in the refugee survey such as mental health (e.g. Kroenke, Baye & Lourens, 2019) or the Big Five (e.g. Danner et al., 2016). Tests for measurement invariance not only validate the questionnaire from a survey methodological point of view, but would contribute to the literature on cross-cultural research in general. The absence of invariance reveals whether, for instance, values or perceptions of mental well-being are universal constructs or whether they are dependent on the cultural setting. Using an explicitly heterogeneous sample such as a sample of refugees is thus a great opportunity to advance the research on this matter.

Regarding the role of institutions as discussed in Chapter 7, a future research project could take the question of safe and unsafe residence titles further and ask how such differences lead to different economic integration in the long run. The dissertation shows that short-term investments in economic integration seem to be negatively affected by a lack of safe residence; however, a perspective on long-term effects is missing. Thus, in a next step, analyses could be conducted on earlier cohorts of refugees to verify whether long-term integration investments decrease as well.

Additionally, an international comparative perspective in this regard is valuable. The way refugees are treated across the world differs greatly. Access to rights and security is not necessarily the normal case. Thus, using the IAB-BAMF-SOEP Survey of Refugees as a basis for cross-country comparisons of how refugees fare in their different host countries (and thus act in different institutions) would be most valuable. Such a research project has already been launched in cooperation with scholars from the University of Manitoba in Canada. In a small project, we have pooled the Canadian Longitudinal Immigration Database (IMDB, StatsCan, 2018) and the IAB-BAMF-SOEP Survey of Refugees. By applying a most different systems design (Anckar, 2008), we are exploring short-term labor market access of refugees in Germany and Canada and thus building on a similar hypothesis to the one I have already developed in this thesis. Preliminary results show that the Canadian institutional framework is more capable of integrating refugees into

the labor market. We argue that this is mostly due to the perspective refugees are provided with as they are granted unlimited residence permits and can already apply for citizenship after three years.

Besides addressing shortcomings of this thesis, future research on the issue of implementing a survey of refugees needs to address challenges that emerge when implementing this survey as a panel study long-term. Therefore, the challenges explored in this thesis are not all-encompassing but rather provide a taste on the challenges to come, when the IAB-BAMF-SOEP Survey of Refugees will be implemented long-term. Determinants of panel-attrition, changes in sensitivity, or interviewer effects are only a few of potential research foci. For example, the third, 2020, release of the IAB-BAMF-SOEP Survey of Refugees reveals that household response in the refugee survey, the second time in a row, is lower compared to the general SOEP population survey (64% and 79% respectively in 2018[2]), and that the research consortiums experiences more quality neutral panel drop-outs as refugees are more likely to move abroad due to return migration or deportation. Considering these differences is a first step towards a deeper understanding on how such a methodologically exotic target population behaves in the setting of a social survey long-term.

The thesis at hand has presented some first challenges and solutions that emerge from implementing a survey of refugees short-term. It describes the process of sampling, discusses languages issues during the field phase, assesses measurement invariance and proposes to consider the refugee specific context when discussing integration trajectories. This thesis is thus a starting point for research that I plan to pursue in the future. Moreover, I hope that this thesis will play a crucial role in expanding our knowledge in the areas of integration research and cross-cultural survey methods, and that future research will see this work as a valuable source and connecting point.

[2] Including quality neutral drop-outs.

References

Ager, A., & Strang, A. (2008). Understanding Integration: A Conceptual Framework. *Journal of Refugee Studies,* 21(2), 166-191. doi:10.1093/jrs/fen016

Anckar, C.. (2008). On the Applicability of the Most Similar Systems Design and the Most Different Systems Design in Comparative Research. *International Journal of Social Research Methodology,* 11(5), 389-401. doi: 10.1080/13645570701401552

Ariely, G., & Davidov, E. (2011). Can we Rate Public Support for Democracy in a Comparable Way? Cross-National Equivalence of Democratic Attitudes in the World Value Survey. *Social Indicators Research,* 104(2), 271-286. doi:10.1007/s11205-010-9693-5

Boehnke, K. (2018). Latent Variable Equivalence: Towards a Quantitative Emic Approach to Cross-Cultural Comparisons. in *OECD-GESIS Seminar on Translating and Adapting Instruments in Large Scale Assessments*. Paris.

Danner, D., Rammstedt, B., Bluemke, M., Treiber, L., Berres, S., Soto, C. J., & Joh, O. P. (2016). Die deutsche Version des Big Five Inventory 2 (BFI-2). https://nbn-resolving.org/urn:nbn:de:0168-ssoar-65715-2, last visited: 02.04.2020

de Maizière, T. (2017). "Wir sind nicht Burka": Innenminister will deutsche Leitkultur. in *Die Zeit*.

Editorial Nature Human Behaviour. (2020). Tell It Like It Is. *Nature Human Behaviour,* 4(1), doi:10.1038/s41562-020-0818-9

Fuchs, J., Kubis, A., & Schneider, L. (2019). Zuwanderung und Digitalisierung. Wie viel Migration aus Drittstaaten benötigt der deutsche Arbeitsmarkt künftig?. Gütersloh: Berterlsmann Stiftung

Hainmueller, J., Hangartner, D., & Pietrantuono, G. (2015). Naturalization Fosters the Long-Term Political Integration of Immigrants. *Proceedings of the National Academy of Sciences,* 112(41), 12651-1256.doi:10.1073/pnas.1418794112

Huntington, S. P. (1993). The Clash of Civilisations? *Foreign Affairs,* 72(3), 22-49. doi:10.2307/20045621

Inglehart, R., & Norris, P. (2003). The True Clash of Civilizations. *Forein Policy,* (135), 62-70.

Jacobsen, J., Eisnecker, P., & Schupp, J. (2017). "In 2016, around one-third of people in Germany donated for refugees and ten percent helped out on site-yet concerns are mountind. *DIW Economic Bulletin,* 16+17, 165-176

Jann, B., Jerke, J., & Krumpal, I. (2011). Asking Sensitive Questions Using the Crosswise Model: An Experimental Survey Measuring Plagiarism. *Public Opinion Quarterly,* 76(1), 32-49. doi:10.2307/41345966

Kroenke, K., Baye, F., & Lourens, S.G. (2019). Comparative Validity and Responsiveness of PHQ-ADS and other Composite Anxiety-Depression Measures. *Journal of Affective Disorders,* (246), 437-443. doi:10.1016/j.jad.2018.12.098

Krosnick, J. A. (1991). Response Strategies for Coping with the Cognitive Demands of Attitude Measures in Surveys. *Applied Cognitive Psychology,* 5(3), 213-36. doi:10.1002/acp.2350050305

Kühne, S.. (2018). From Strangers to Acquaintances? Interviewer Continuity and Socially Desirable Responses in Panel Surveys. *Survey Research Methods,* 12(2), 121-46. doi:10.18148/srm/2018.v12i2.7299

Nederhof, A. J. (1985). Methods of Coping with Social Desirability Bias: A Review. *European Journal of Social Psychology,* 15(3), 263-280. doi:10.1002/ejsp.2420150303

Parsons, T. (1968). *The Structure of Social Action*. New York: Free Press.

Perry, A,, & Gauly, B. (2019). Analysing Poor Reading Skills: A Comparison Between PIAAC Literacy Levels and Reading Components. *Survey Methods: Insights from the Field*. doi:10.13094/SMIF-2019-00020

Schwartz, S. H., & Sagie, G. (2000). Value Consensus and Importance: A Cross-National Study. *Journal of Cross-Cultural Psychology,* 31(4), 465-497. doi:10.1177/0022022100031004003

StatsCan. (2018). Longitudinal Immigration Database (IMDB) Technical Report, 2018. Statistics Canada.

Tourangeau, R., & Yan, T. (2007). Sensitive Questions in Surveys. *Psychological Bulletin,* 133(5), 859-883. doi:10.1037/0033-2909.133.5.859

Dissertation Appendix

1) Statement of Authorship

2) Availability of Syntax

Statement of Authorship

Dissertation "Establishing a Survey of Refugees in Germany: Challenges in Sampling, Field Work and Measurement".

I expressly declare that the work I have submitted was written independently and without external help.

I expressly declare that all sources used in the abovementioned work – including those from the Internet (including tables, graphic and suchlike) – have been marked as such. In particular, I declare that I have, without exception, stated the source for any statements quoted verbatim and/or unmodified tables, graphics etc. (i.e. quotations) of other authors.

I am aware that violations against the principles of academic independence are considered deception and are punished accordingly.

Jannes Jacobsen, Berlin April 3rd 2020

Availability of Syntax

All work that is part of this thesis was carried out using Stata 14 SE and RStudio. The Syntax is available either upon request or online in my personal GitHub repository at https://git.soep.de/jjacobsen/dissertation-jacobsen